PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/10, 15/12, 15/24, 15/62, 15/70

A1

(11) International Publication Number: **WO 97/32017**

(43) International Publication Date: 4 September 1997 (04.09.97)

(21) International Application Number: PCT/EP97/00931

(22) International Filing Date: 26 February 1997 (26.02.97)

(30) Priority Data:
96102852.9 26 February 1996 (26.02.96) EP

(34) Countries for which the regional or international application was filed: DE et al.

(71) Applicant (*for all designated States except US*): MORPHOSYS GESELLSCHAFT FÜR PROTEINOPTIMIERUNG MBH [DE/DE]; Frankfurter Ring 193a, D-80807 München (DE).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): ILAG, Vic [PH/DE]; Knorrstrasse 85, D-80807 München (DE). GE, Liming [CN/DE]; Portiastrasse 12, D-81545 München (DE).

(74) Agent: VOSSIUS & PARTNER; P.O. Box 86 07 67, D-81634 München (DE).

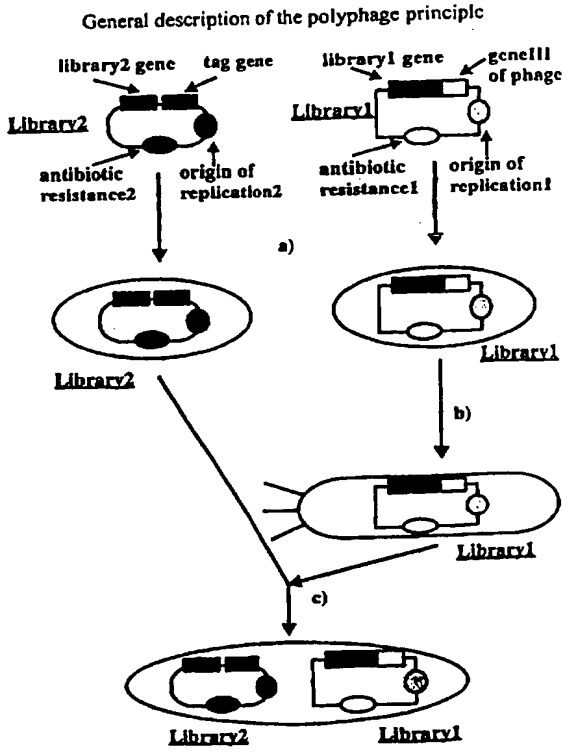
(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NOVEL METHOD FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING TWO OR MORE INTERACTING (POLY)PEPTIDES

(57) Abstract

The present invention relates to methods for identifying nucleic acid sequences which encode two or more specific interacting peptides or proteins. Furthermore, the present invention relates to kits which may be used for the identification of nucleic acid sequences in accordance with the method of the present invention.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

NOVEL METHOD FOR THE IDENTIFICATION OF NUCLEIC ACID SEQUENCES ENCODING TWO OR MORE INTERACTING (POLY)PEPTIDES

The present invention relates to methods for identifying nucleic acid sequences which encode two or more specific interacting peptides or proteins. Furthermore, the present invention relates to kits which may be used for the identification of nucleic acid sequences in accordance with the method of the present invention.

Protein-protein interactions play an important role in all biological processes, from the replication and expression of genes to the morphogenesis of organisms (Lewin, B. 1994, Genes V. Oxford University Press). Methods for detecting protein-protein interactions have proved useful in understanding the basic mechanisms of different biological processes and the development of therapeutics. Detection of protein-protein interactions can be divided into two main categories: (i) physico-chemical based and (ii) genetic approaches (Phizicky, E.,M. & Fields, S. Microbiological Reviews 59 (1995) 94-123). Detection of protein-protein interactions by physico-chemical methods usually requires significant amounts of material, and more importantly, the identity of the proteins to be studied must be known. Recent developments in methods of mass spectrometry circumvent this problem but such suffer the disadvantage of requiring sophisticated equipment and expertise (Wang, R. & Chait, B.T., Current Opinion in Biotech. 5 (1994) 77-84). In contrast, genetic approaches provide an easy and powerful method of identifying protein-protein interactions without the need for pure material and specialized equipment, with the added advantage of higher throughput.

Different genetic approaches have been used to identify protein-protein interactions. The current method of choice is the yeast 2-hybrid system (Fields, S. & Song, O.K.,

Nature (London) 340, (1989) 245-246) which allows the identification of novel proteins that interact with a known protein.

Another popular genetic approach is the phage display system (Patent Application WO90/02809) whereby proteins are fused to a component of a surface protein of filamentous phage to allow selection for binding to a ligand of interest. The gene encoding the protein displayed on the surface of the phage is packaged inside the phage allowing the coupling of genetic information with the gene product. This allows the screening of "libraries" of proteins whereby the identity of the screened protein is deduced from the nucleic acid sequence of the phage. This technique has been extended by Winter et al. (Patent Application WO 92/20791) to produce libraries of multimeric members of a specific binding pair (e.g. combinations of VH and VL chains of an antibody) and select for functional specific binding pair members that can bind to the complementary specific binding pair member (e.g. antigen). Said libraries are constructed by combining two sub-libraries each encoding a collection of corresponding sub-units of said multimeric members (e.g. a library of VH chains is combined with a library of VL chains) wherein in principle each sub-unit out of the first sub-library is able to bind to each sub-unit out of the second sub-library non-specifically. Although this method has led to the identification of unique antibodies against particular antigens, it fails to provide a method for identifying two partners of a specific binding pair when both are unknown.

A unique version of phage display which relies on non-infective phage has recently been proposed (Duenas, M. & Borrebaeck, C. A. K., Bio/Technology 12 (1994) 999-1002; EP 0 614 989). A version of this system that led to the identification of proteins from a cDNA library that interacts with the jun protein has been described (Gramatikoff et al., Nucleic. Acids Res. 22 (1994) 5761-5762). The same principle has been also shown to work with an antibody-antigen system (Krebber et al., FEBS Letters 377 (1995) 227-231).

In spite of the power of all the aforementioned genetic selection approaches, they are limited to the selection of interacting binding entities from only a single genetically-diverse population (library vs. individual).

It would, however, be highly desirable to simultaneously identify binding entities and their specific binding partners in a library vs. library setting, wherein preferably at least two genetically diverse populations are involved. A solution to this technical problem, i.e. the identification of interacting entities and the respective nucleic acid sequences from more than one genetically diverse population (library vs. library) is neither provided nor suggested by the prior art. The present invention solves the above technical problem by providing the embodiments characterized in the claims. By using these embodiments, it has become possible to increase exponentially the rate at which (poly)peptide-(poly)peptide interactions are detected. The present invention may find applications in the field of functional genomics, whereby different proteins of unknown functions can be related with other proteins.

Accordingly, the present invention relates to a method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

(a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

(b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (a), wherein the vector molecules

employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (a) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (a) and (b), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

(c) optionally, providing additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (a) and/or step (b), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (a) and (b) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (c) together with at least one of said properties displayed by either said vector molecule and/or said recombinant insert used in steps (a) and/or (b), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

(d) expressing members of said libraries of recombinant vectors or nucleic acid sequences mentioned in steps (a), (b) and optionally (c), in appropriate host cells so that at least one interaction is established;

(e) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;

(f) optionally, carrying out further selection, screening and/or purification steps; and

(g) identifying said nucleic acid sequences encoding said (poly)peptides.

Thus, in the context of the present invention, the term "properties that are phenotypically distinguishable" relates alternatively to properties that are encoded by the vector molecule or to properties that are encoded by the recombinant insert or to both types of properties. As regards the vector-encoded properties, these may e.g. be resistance markers or requirements for special nutrients. It should be noted that the recombinant insert may comprise a nucleic acid portion encoding said property in addition to the nucleic acid portion responsible for the interaction.

In the context of the present invention, the term "different member " denotes a different entity which may be, but is not necessarily, structurally different.

Further, in the context of the present invention, the term "plurality" bears the meaning of "at least two".

The novel properties generated by the at least two recombinant inserts reflect the inventive principle of the present invention. That is, only if two (or more) (poly)peptides interact, for example, in a homo-dimeric or hetero-dimeric fashion, a screenable or selectable property is generated. The interaction between the two or more molecules may be a direct one or may be mediated indirectly. Examples for a direct interaction are the binding of an antibody encoded by a nucleic acid sequence from library 1 to a cDNA protein from library 2, the binding of a protein encoded by a nucleic acid sequence from cDNA library 1 to a protein from a cDNA library 2, as well as of an anti-idiotypic antibody encoded by a nucleic acid sequence from one of the libraries to a corresponding antibody encoded by a nucleic acid sequence from the other library. The nucleic acid sequences are preferably DNA and most preferably genes or parts thereof.

An example of an indirect interaction is the bridging of two (poly)peptides encoded by the two libraries which is mediated by a phosphorylating enzyme. Once the phosphorylation of one (poly)peptide encoded e.g. by library 1 is effected by the respective kinase, then this protein is capable of interacting with the second (poly)peptide encoded by library 2. The phosphorylating enzyme exemplifying this type of interaction may be encoded by a nucleic acid from (one of) the additional libraries and/or may be encoded by the genome of the host cell. Typically, the interaction of the two (poly)peptides forms a "bridge" of molecules, said "bridge" being detectable using an appropriate detection process. Conveniently, said bridge is detectable by a tag molecule that is associated with, encoded by or attached to one of the (poly)peptides encoded by library 1 or preferably 2.

Furthermore, the present invention relates to a method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

(a) expressing in appropriate host cells

(aa) nucleic acid sequences contained in a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

(ab) nucleic acid sequences contained in a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as

mentioned in step (aa), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (aa) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (aa) and (ab), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

(ac) optionally, nucleic acid sequences contained in additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (aa) and/or step (ab), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (aa) and (ab) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (ac) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (aa) and/or (ab), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

so that at least one interaction is established;

(b) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;

(c) optionally, carrying out further screening, selection and/or purification steps; and

(d) identifying said nucleic acid sequences encoding said (poly)peptides.

In a preferred embodiment of the method of the present invention, said screenable or selectable property is expressed extracellularly.

This embodiment is conveniently employed in a number of laboratories which would make use of rather conventional methodology of the extracellular detection of such properties, e.g. by column chromatography wherein the e.g. screenable tag is retained, in combination with e.g. plaque purification techniques, which allow the further purification of the cells that were originally enriched by e.g. the column chromatography step.

In a further preferred embodiment of the method of the present invention, said recombinant vector molecule in step (a)/(aa) (the step identified after the slash refers to the corresponding step of the second embodiment of the method of the invention identified hereinabove) gives rise to a replicable genetic package (RGP) displaying said (poly)peptides at its surface. In this context, the term replicable genetic package (RGP) refers to an entity, such as a virus or bacteriophage, which can be replicated following infection of a suitable host cell. In the case of bacteriophage, for example, the collection of nucleic acid sequences can be inserted into either a phage or phagemid vector in frame with a component of the phage coat, such as gene III, resulting in display of the encoded binding entities on the surface of the phage. Particularly preferred as a

recombinant vector molecule is a recombinant phage, phagemid or virus, wherein said phage is most preferably

(a) one of the class I phage fd, M13, If, Ike, ZJ/2, Ff;

(b) one of the class II phage Xf, Pf1, and Pf3;

(c) one of the lambdoid phages, lamda, 434, P1;

(d) one of the class of enveloped phages, PRD1; or

(e) one of the class paramyxo-viruses, orthomyxo-viruses, baculo-viruses, retro-viruses, reo-viruses and alpha-viruses.

In a further preferred embodiment of the method according to the invention, said selection step (e)/(b) is carried out by selecting polyphage comprising the interacting (poly)peptides. Polyphage contain more than one copy of phage genomic DNA. They occur naturally at a low to moderate frequency when a newly forming phage coat encapsulates two or more single-stranded DNA molecules. In the case of the present invention, the polyphage which are formed will contain at least two phage genomes, which may either (i) both be representatives of library 1, or (ii) both be representatives of library 2, or (iii) be representatives of each of library 1 and library 2, or (iv) be a combination of (i) to (iii) with at least one member of one of the additional libraries. The efficiency of polyphage production can be increased by the introduction of appropriate mutations into the phage genome, as is well known to those skilled in the art (see, for example, Lopez, J. and Webster, R.E., Virology 127 (1983), 177-193, Bauer, M. and Smith, G.P., Virology 167 (1988) 166-175, or Gailus, V. et al., Res. Microbiol. 145 (1994) 699-709).

In a further preferred embodiment of the method of the invention, said screenable or selectable property is connected to the infectivity of said RGP.

In this embodiment, use is made of the possibility that the infectivity of e.g. a bacteriophage can be manipulated, said infectivity being directly correlated with the interaction of said (poly)peptides.

In a most preferred embodiment of the method of the present invention, said RGP is encoded by said recombinant vector used in step (a)/(aa) and rendered non-infective and infectivity of said RGP is restored by interaction of said (poly)peptide of step (a)/(aa) with the (poly)peptide of step (b)/(ab) and/or (c)/(ac), said (poly)peptide of step (b)/(ab) and/or (c)/(ac) being fused to a domain that confers infectivity to said RGP.

In a further most preferred embodiment of the method of the invention, said RGP is rendered non-infective by modification of a genetic sequence which encodes a surface protein necessary for the RGP's binding to and infection of a host cell.

These preferred and most preferred embodiments of the method of the present invention relating to the infectivity of the RGP serve as an alternative to the use of the screenable tag. In these embodiments, advantage can be taken of the phenomenon of selective infection (Krebber et al., FEBS Letters 377 (1995) 227-239). While the screenable tag enables physical separation of molecules from others in the population, the use of selective infection enables positive selection for the interacting pair. This phenomenon relies on the use of a construct which can selectively restore infectivity to phage which have been rendered non-infective by, for example, deletion of all but the C-terminus of the gene III protein. Use of such phage for displaying library 1 gives non-infectious phage carrying the binding entity. Co-expression with library 2 allows interactions between binding entities and binding partners to be established, as described above. Although the phage which carry the binding entity-binding partner pair are non-infective, infectivity can be restored if, in place of the screenable tag referred to above, an infectivity protein is used. In this context, the term infectivity protein refers to a substance which, when associated with the phage, can enable it to penetrate a bacterial host, where it is subsequently replicated. An example of an infectivity protein is the N-terminus (at least the first 220 amino acids) of gene III protein of the filamentous bacteriophage.

The infectivity protein confers on those phage which carry it, the ability to be replicated. Thus, only those phage which carry the binding entity/partner pair are replicated. Purification of hybrid phage containing genes from both libraries 1 and 2 then relies e.g. on the use of two selectable markers as indicated above. The genes in the phage can then be identified using methodology well known to those skilled in the art.

An additional preferred embodiment of the present invention relates to a method, wherein said recombinant vector molecules in step (a)/(aa) give rise to a fusion protein which is expressed on the surface of a cell, preferably a bacterium.

These fusion proteins, upon interaction with a suitable binding partner from library 2 connected e.g. with a screenable tag can be detected on the surface of host cells which may be, for example, bacteria, yeast, insect cells or mammalian cells. The display of fusion proteins on bacterial surfaces per se is well known in the art. Thus, lipoproteins (Lpp), outer membrane proteins A (OmpA), and flagella have been used to target antibodies and peptides to the cell surface of E.coli. Fuchs et al., Bio/Technology 9 (1991) 1369-1372, WO93/01287, presented a single chain antibody on the surface of E.coli as a fusion protein with the N-terminus of the peptidoglycan-associated lipoprotein. The antibody was visualized by the binding of fluorescently labeled antigen and fluorescently labeled antibodies directed to the linker peptide of the displayed single chain antibody. Francisco et al., Proc. Natl. Acad. Sci. USA 90 (1993) 10444-10448, and Georgiu, G. et al., WO93/10214, displayed antibodies on the E.coli surface by fusing the N-terminus of a single chain antibody to the C-terminus of OmpA while the N-terminus of OmpA was fused to the signal sequence and the first nine amino acids of Lpp. Binding of a fluorescently labeled antigen to the OmpA-antibody fusion protein was detected by FACS. Klauser (WO 95/17509) transferred the IgA protease system from Neisseria to E.coli to facilitate display of antibodies. Integration of the beta-domain of the IgA protease precursor into the outer membrane lead to the transport of the

protease domain across the membrane followed by autoproteolytic release into the medium. Antibodies linked to the beta-domain of IgA protease are therefore presented on the surface of bacteria. Further, Lu, Z. et al., Bio/Technology 13 (1994) 366-371, described a system for displaying peptides on the surface of the bacterium by fusing it to thioredoxin and the bacterial flagella, to screen for peptide mimics of the epitope for an anti-IL-8 antibody.

The further identification of the desired nucleic acid molecule encoding the interacting (poly)peptides may then be effected by methods known in the art, e.g. by purifying host cells displaying a tag on their surface and further by antibioticum-based selection techniques, DNA purification and sequencing.

In a particularly preferred embodiment of the method of the present invention, said bacterium is Neisseria gonorrhoe or E.coli and said fusion protein consists of at least a part of a flagellum, lam B, peptidoglycan-associated lipoprotein or the Omp A protein and said (poly)peptide.

As has been repeatedly pointed out hereinabove, a tag connected to the (poly)peptide encoded by library 2 can conveniently be used in the identification strategy of the desired nucleic acid sequences. Accordingly, in a further preferred embodiment of the method of the invention, said (poly)peptides encoded by said recombinant vector molecules of step (b)/(ab) or (c)/(ac) are linked to at least one screenable or selectable tag. In this context, the term screenable or selectable tag refers to a short sequence of amino acids which can be recognized and bound by a particular substance. Tags are commonly used for the purification of biomolecules: examples are His(n), where n = 4-6 which can be bound either by Ni, or a specific antibody, and the flag and myc tags which are recognized by appropriate antibodies. In either of these cases, the tag can be encoded as a C-terminal fusion to all binding partners in library 2. In accordance with the present invention, the tag can be used to isolate e.g. the polyphage referred to

above. Thus, the interaction between the phage-bound binding entity, and its interacting binding partner, establishes a connection between the phage particle and the screenable or selectable tag. This feature can be exploited in a step which relies on e.g. affinity chromatography to isolate the polyphage carrying the interacting molecules. In a final step, those polyphage which carry two distinct nucleic acid molecules and preferably genes (encoding binding entity and binding partner) can be separated from those carrying only one of the two genes e.g. by selection based on transduction or different selectable markers (e.g. antibiotic resistance) present in the individual genomes. In this way, the genes which encode the two interacting molecules can be identified.

A most preferred embodiment of the present invention relates to a method wherein said screenable or selectable tag is encoded by said recombinant vector of step (b)/(ab) or (c)/(ac).

A further most preferred embodiment of the present invention relates to a method wherein said screenable or selectable tag is selected from the list His(n), myc, FLAG, malE, thioredoxin, GST, streptavidin, beta-galactosidase, alkaline phosphatase T7 gene 10, Strep-tag and calmodulin. These screenable tags are all well known in the art and are fully available to the person skilled in the art.

In an additional particularly preferred embodiment of the method of the invention, said screenable or selectable tag is encoded by the genome of the host cell.
An example for this embodiment is an anti-Fc-receptor specific antibody that is expressed by the host cell and could function as an additional bridge in e.g. purification by column chromatography. Another example of this embodiment is an enzyme produced by the host cell that creates a tag such as a phosphorylation on (poly)peptides of the second library without destroying the interaction of (poly)peptides of step (b)/(ab)

with (a)/(aa) so that the modification caused by the enzyme is now the screenable or selectable tag.

In a further preferred embodiment of the method of the invention, said (poly)peptides encoded by the nucleic acid sequences of said additional libraries of step (c)/(ac) cause the interaction of said (poly)peptides of steps (a)/(aa) and (b)/(ab) via phosphorylation, glycosylation, methylation, lipidation or farnesylation of at least one of said (poly)peptides of steps (a)/(aa) and (b)/(ab).

An additional preferred embodiment of the invention relates to a method wherein said host cells in step (d)/(a) are spatially addressable, and the nucleic acid sequences mentioned in step (g)/(d) are retrieved from the corresponding spatially addressable host cell.

In the context of the present invention, the term "spatially addressable" refers to a situation where the individual cells harboring one of the potential combinations of members of the first, second and optionally additional libraries are identifiable by their relative position, e.g. by their position on a master plate. The screening or selection may, for example, be performed either with single clones derived from the master plate, or on a replica plate, thus maintaining the connection between the screenable or selectable property and the information contained in the host cell on the master plate.

An additional preferred embodiment of the invention relates to a method wherein said screenable or selectable property is expressed intracellularly.

Particularly preferred is a method wherein said screenable property is the transactivation of the transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu or resistance genes

giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline, or streptomycin.

Furthermore, use can be made of the yeast 2-hybrid system referred to hereinabove or the interaction trap system (Brent et al., EP-A 0 672 131) or of a prokaryotic version analogous to the above recited systems, utilizing the toxR system of Vibrio cholerae (Fritz, H.-J. et al., EP-A 0 630 968). It is within the skills of the person skilled in the art to combine further screening systems known in the art with the method of the present invention.

In a further preferred method of the present invention, said recombinant vectors of step (a)/(aa), (b)/(ab) and (c)/(ac) comprise recombination promoting sites and in said step (e)/(b) recombination events are selected for, wherein said nucleic sequences encoding said (poly)peptides of step (a)/(aa), said nucleic acid sequences encoding said (poly)peptides of step (b)/(ab) and optionally said nucleic acid sequences encoding said (poly)peptides of step (c)/(ac) are contained in the same vector. In this approach, the two genes can be coupled in a single vector, and packaged in a phage of standard size, if appropriate recombination sites are incorporated in the vectors carrying libraries 1 and 2. Again, the phage which carry both nucleic acid sequences and genes are purified with the use of e.g. the screenable tag. If recombination is used to couple the genes from the two libraries, some of the hybrid progeny phage will contain nonrecombinant genomes, since site-specific recombination is not very efficient. However, the hybrid phage can be selected by re-infection of host cells that do not contain library 2 followed by another round of selection of the screenable tag.

In a particularly preferred embodiment of the method of the invention, said recombination events are mediated by the site-specific recombination mechanisms Cre-lox, attP-attB, Mu gin or yeast flp.

In a further particularly preferred embodiment of the method of the invention, said recombination promoting sites are restriction enzyme recognition sites and said recombination event is achieved by cutting the recombinant vector molecules mentioned in steps (a)/(aa), (b)/(ab) and optionally (c)/(ac) with at least two different restriction enzymes and effecting recombination of the nucleic acid sequences contained in said vectors by ligation.

The invention relates in an additional preferred embodiment to a method wherein said identification of said nucleic acid sequences is effected after the selection step (e)/(b) via PCR and preferably sequencing of said nucleic acid sequences after said PCR.
After said selection step (e)/(b), PCR can be carried out with the enriched desired product, conveniently using primers that hybridize to the vector portion of the recombinant vector molecule. Sequencing of the PCR-product may then be carried out according to conventional methods.

In a further preferred embodiment of the method according to the invention, said recombinant vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac) comprise at least one gene encoding a selection marker.
Said genes encoding said selection markers are preferably different in each of the vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac), i.e. said vectors comprise genes encoding different selection markers. Said selection markers can conveniently be used for the further purification envisaged in step (f)/(c). For example, a polyphage comprising two members of each library 1 and 2 can be selected for on the basis of a double resistance to antibiotics. Also, a successful recombination event may create a new recombinant vector carrying both nucleic acid molecules from library 1 and 2 as well as genes encoding different selection markers. Again, the selection for a twofold resistance will assist in the identification of the desired product.

In a particularly preferred embodiment of said method, said selection marker is a resistance to an antibiotic, preferably to ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

A further preferred embodiment of the present invention relates to a method wherein said host cells are F' and preferably E.coli XL-1 Blue, K91 or its derivatives, TG1, XL1kan or TOP10F.

In a particularly preferred embodiment of the present invention, said RGPs are produced with the use of helper phage taken from the list R408, M13k07 and VCSM13, M13de13, fCA55 and fKN16 or derivatives thereof.

Further preferred is a method wherein at least one of said genetically diverse nucleic acid sequences encode members of the immunoglobulin superfamily.

Said method is particularly preferred, if said genetically diverse nucleic acid sequences encode a repertoire of immunoglobulin heavy or light chains.

In an additional preferred embodiment of the present invention, in said method said genetically diverse nucleic acid sequences are generated by a mutagenesis method. Various mutagenesis methods are well known to the person skilled in the art and need not be described in here in any further detail.

The present invention relates in an additional preferred embodiment to a method in which said genetically diverse nucleic acid sequences are generated from a cDNA library.

In a final preferred embodiment of the method of the invention, said nucleic acid sequences are genes or parts thereof.

As used herein, the term "parts thereof" relates to parts of genes that encode a product that is capable of interacting with a product encoded by any of the other libraries. Thus, it is well known that various proteins are comprised of different domains. Only one of said domains may be capable of interacting with a different (poly)peptide. Such a domain might be encoded by a part of said gene in accordance with the present invention.

The invention also provides for identifying genes encoding more than two interacting peptides or proteins. This can be achieved by using additional vectors encoding genetically diverse additional nucleic acids by an extension of the method described above. As previously, the presence of either a screenable tag or an infectivity protein is used to purify phage carrying genes which encode the components of the complex. Again, the genes in the phage can then be sequenced using methodology well known to those skilled in the art.

Additionally, the present invention relates to a kit comprising at least

(a) a recombinant vector molecule as described in step (a)/(aa) or a corresponding vector molecule;

(b) a recombinant vector molecule as described in step (b)/(ab) or a corresponding vector molecule; and, optionally,

(c) at least one further recombinant vector molecule as described in step (c)/(ac) or a corresponding vector molecule.

As a rule, if recombinant vector molecules are comprised in said kit, they will comprise a library of nucleic acid molecules. In other words, the kit of the invention will contain a plurality of different recombinant vector molecules.

Legends to Figures and Tables

Figure 1: General description of the polyphage principle

a) transform to E. coli hosts

b) infect host containing library1 with helper-phage to package library1 into phage

c) infect cells containing library2 with phages containing library1 leading to cells harboring members of library1 and library2; the presence of library1 and library2 is selected by the presence of the 2 antibiotic resistance markers

d) expression of library1 and library2-tag gene products

e) infect cells with engineered helper-phage to induce polyphage production

Note 1: Polyphage does not discriminate which genome to package therefore the possibilities resulting from step e) arise in an infected cell. To select for the polyphage containing the right packaged genomes the subsequent step is required

f) select for tag e.g., infectivity-mediating protein, in which case ability to infect is selected and

g) select for ability to confer resistance to 2 antibiotics to infected cells

Note 2: Only polyphages that satisfy f) + g) represent phages that display the correct interacting pair and the corresponding genetic information

Figure 2: Co-transformation of two phagemids, polyphage formation and selection *via* His-tag: general description

A, B: libraries of phagemids, preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); after co-transformation phage production leading to a phage population displaying cognate pairs (left part of the Figure) or not (right part), after selection infection of host cells, selection for double-resistance

Alternative methods include the infection of cells harbouring a plasmid- or phagemid-based library B with a phage library A (prerequisite again: interference-resistant constructs).

Figure 3: pBS vector series: functional map and sequence of pBS13

Figure 4: Co-existence of phagemids: results of restriction digest
Restriction analysis of clones of double resistances (Amp/Cm). R1: pIG10.3, *Xba/ScaI*; R2: pBS13, *Xba/ScaI*, R1+R2: R1 and R2 are mixed in approx. equal proportion; M1: marker λ: *Bst*EII; M2: marker pBR322: *Msp*I; 1 to 10: randomly picked clones: *Xba/ScaI*

Figure 5: Phagemid vector pYING1-C1: functional map
containing the fos peptide. The corresponding vectors pYING1-C2 and pYING1-C3 contain instead of fos the p75 and the IL16 peptides, respectively

Figure 6: Phagemid vector pYANG3-A: functional map
containing the jun peptide. The corresponding vectors pYANG3-Ape2, pYANG3-Ape3, and pYANG3-Ape10 contain instead of jun the p75-binding peptides pe2, pe3, and pe10, respectively

Figure 7: Analysis of selected clones (see Table 2):
7.a: Restriction digest of clones before and after selection
R: pYANG3-Ape2: *Xba*I; M1: marker λ: *Bst*EII; M2: marker pBR322: *Msp*I; α/1 to 10: randomly picked clones before selection: *XbaI/Hin*dIII; β/1 to 10: randomly picked clones after selection: *XbaI/Hin*dIII; size expected: jun-gIII: 745 bp; fos: 256 bp; p75: 577 bp; IL-16: 502 bp
7.b: PCR reaction of clones after selection with primers OPEP5L and OGIII3
R1: pYANG3-A as template; R2: pYANG3-Ape2 as template; M: marker λ: *Bst*EII; β/1 to 10: randomly picked clones after selection as templates

Figure 8: Phagemid vector pING1-C1: functional map
containing the His-tag peptide. The corresponding vector pING3-C1 contains an additional FLAG epitope; pING1-C2 and pING3-C2 contain

the Strep-tag instead of His-tag, with pING3-C2 containing an additional FLAG epitope.

Figure 9: Phagemid vector pONG3-A: functional map
for the generation of phage-display libraries (gIII fusions)

Figure 10: Co-transformation of phage and plasmid, polyphage formation and selection *via* SIP: general description
fA: library A in phage construct; B: library B, library members fused to IMP; preferably different resistance markers on phage and plasmid; after co-transformation production of phages; in the case of cognate-pair interaction formation of infectious phages; selection; by plating on double-resistance identification of polyphage particles.

Figure 11: Phage vector fhag1A: functional map
for phage-display of the α-HAG scFv

Figure 11a: CAT gene module: functional map and sequence

Figure 12: Phage vector fjun1A: functional map
for phage-display of the jun peptide

Figure 13: Phage vector fjun1B: functional map
for phage-display of the jun peptide

Figure 14: Phage vector fpep3_1B: functional map
for phage-display of the peptide pe3 binding to the intracellular domain of p75

Figure 15: Phage vector fNGF_1B: functional map
for phage-display of NGF

Figure 16: Plasmid pUC19/IMPhag: functional map
containing fusion of HAG peptide to the N-terminal domains of gIIIp (IMP)

Figure 17: Plasmid pUC18/IMPp75: functional map
containing fusion of the intracellular domain of p75 to the N-terminal domains of gIIIp (IMP); pUC18/IMPfos contains the fos peptide instead of the intracellular domain of p75

Figure 18: Plasmid pUC18/IMPIL16: functional map containing fusion of IL16 to the N-terminal domains of gIIIp (IMP)

Figure 19: Analysis of selected clones (see Table 3)

Lane 1: marker λ: *Bst*EII; lanes 2 to 20: polyphage transductant clones #1 to #19 digested with *Xba/Hind*III; f.._1b: fragment of phage vector after digest; pUC18: fragment of plasmid after digest; α-HAG: fragment containing anti-HAG scFv fused to gIIIc; IMP-p75 and IMP-HAG: fragment containing IMP fused to p75, and IMP-HAG peptide, respectively; pep3-gIIIs: fragment containing pep3 fused to gIIIc (s: short version)

Figure 20: Co-transformation of phagemids, *in vivo* recombination and selection *via* His-tag: general description

A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); both constructs containing recombination-promoting sites (*) such as lox/loxP; after co-transformation and recombination production of phages; selection *via* Ni-NTA; re-infection of host cells, selection for double-resistance

Figure 21: *In vitro* recombination and selection *via* His-tag: general description

A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to gIIIp; B: fusions to tag (His); both constructs containing corresponding recognition sites for restriction enzymes (+/o); after digest and co-ligation transformation and production of phages; selection *via* Ni-NTA; re-infection of host cells, selection for double-resistance

Figure 22: Phage vector fjunhag: functional map for phage display of the jun peptide

Figure 23: Spatial *in vivo* SIP: general description

After transformation or co-transformation according to any of the methods described above, a master plate is made. From that phages secreted from individual clones can be analyzed individually (top), or a replica (migration of secreted phages through filter disc) can be made whereon selection for the presence of a tag or infectivity can be performed. By going back to the

master-plate, the information for selected cognate interacting pairs can be retrieved without requiring recombination and/or polyphage production.

Figure 24: *E. coli* display: general description

A, B: libraries of phagemids; preferably with different resistance markers; A: fusions to E.coli surface-display protein; B: fusions to tag (His); after co-transformation expression of constructs; surface-display; in the case of cognate interaction taking place, display of tag on the surface of the host cell; selection

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence

Table 1: Phagemids constructed for Experiments 2 and 3

Table 2: Results of Experiment 2 (see Figure 7)

2.a: Combination of phagemids present in initial library (α)

2.b: Combination of phagemids present after selection (β)

Table 3: Results of Experiment 4 (see Figure 19)

3.a: Identification of phage/plasmid present in individual clones

3.b: Test for infectivity of individual clones

The examples illustrate the invention.

Example 1: General description of th polyphage principl (Figure 1)

The binding entities which comprise library 1 may be peptides or proteins, and are encoded by a genetically diverse collection of first nucleic acid sequences. These nucleic acid sequences are inserted into a first vector which allows for display of the encoded binding entities on the surface of a replicable genetic package. For the purposes of subsequent selection, the first vector should also carry a gene encoding a selectable marker, such as an antibiotic resistance. The binding partners which comprise library 2 may be peptides or proteins, and are encoded by a genetically diverse collection of second nucleic acid sequences which are inserted into a second vector. By way of example, this second vector may be a plasmid, or even a phage or phagemid, in which case the origin of replication should be distinct from that of the first vector. For the purposes of subsequent selection, the second vector should also carry a gene encoding a selectable marker, such as an antibiotic resistance, preferably distinct from that present in the first vector. To facilitate purification of the complex to be formed between any binding entity-binding partner pair, a screenable tag can be conveniently attached to members of library 2.

The two genetically diverse collections of nucleic acids are then introduced into a population of host cells in such a way that encoded libraries 1 and 2 can be expressed. This can be achieved by either (i) co-transformation of the two vectors, or, as actually shown in the figure, (ii) packaging one of the collections of nucleic acids into a vector (such as a bacteriophage) which can be used to infect with high efficiency a population of cells into which the complementary collection of nucleic acid has been introduced. The result is a population of cells in which individual cells carry representatives of each library.

Expression of the two collections of nucleic acids results in the production of pairs of molecules, one from each library, in the host cells. In each case, one or more members

of the library of binding entities is incorporated into the coat of an RGP. In some cells, an interaction will be established between a binding partner on the surface of the RGP and a binding partner expressed from library 2. When such an interaction is established, the RGP therefore carries both the binding entity and the binding partner.

The RGPs displaying such an interaction can then be further purified with the help of polyphage and differing selection markers, as has been discussed hereinabove. After such selection, the nucleic acid sequences encoding one or both binding partners can be conveniently identified by methodology known in the art, such as DNA sequencing.

Example 2: Co-transformation of phagemids with same *E. coli* origin of replication, polyphage formation, and selection of correct pairing interactions *via* His-tag

2.1: Principle (see Figure 2)

To demonstrate that polyphage formation allows the retrieval of the genetic information for cognate protein pairs selected using a tag fused to one member of the protein pair, two separate, small libraries in phagemid vectors are constructed.

2.2: Test of co-existence of phagemids with the same *E. coli* origin of replication:

Prerequisite for the formation of polyphage particles containing two different phagemids is that the different phagemid vectors can co-exist in the host cell.

The vector pBS13 is a derivative of the vector (Krebber *et al.*, 1996) containing a chloramphenicol-resistance gene instead of the kanamycin-resistance gene and a beta-lactamase gene cassette instead of the 2H10-gIII fusion gene, and can be assembled by standard methods starting from pto2H10a3s. Figure 3 contains the functional map and the sequence of pBS13. pIGHAG1A (see Example 4.2.1.f) is digested with *Xba*I and *Hin*dIII. The 1.3 kb fragment containing the anti-HAG gene fused with the C-

terminal domain of filamentous phage pIII protein is isolated and ligated with a pre-digested phagemid vectors pIG10.3, and pBS13 (XbaI-HindIII) to create the vectors pIG10.3-scFv(anti-HAG) (Ap^R) and pBS13-scFv(anti-HAG) (Cm^R), respectively. The vectors are used to transform competent XL-1 Blue cells and selected on LB plates containing Amp/Cm/Tet and glucose (20 mM).

The phagemids from clones of double-resistant colonies (Amp/Cm) are isolated. The restriction digestions indicate the co-isolation of both phagemids from the single colonies (Figure 4).

2.3: Design of libraries A and B:

Library A contains three cyclic peptides each binding to the intracellular domain of the low affinity nerve growth factor (NGF) receptor (see Example 4), and a leucine zipper domain derived from the jun transcription factor, all N-terminally fused to the C-terminal domain of gIII from filamentous phage.

Library B encodes 3 members, namely the leucine zipper domain of the fos transcription factor which heterodimerizes with jun *via* this domain, the intracellular domain of the NGF receptor p75, and, as a negative control which does not interact with library A members, IL-16, all fused at the N-terminus with a His_6-peptide as tag (Hochuli *et al.*, 1988; Lindner *et al.*, 1992).

The cognate pairings are from the interaction between jun and fos (Crameri and Suter, 1993), and p75 and selected cyclic peptides (see Example 4). A non-cognate pairing would occur among the non-cognate pairs mentioned and among jun, or one of the cyclic peptides, and IL-16.

2.4: PCR amplification of the individual constructs

Fos, N-terminus fused to His_6, is PCR amplified using **pOK1** (Gramatikoff *et al.*, 1994) as template and oligonucleotides OFOS-5 and OFOS-3 as primers, where His_6 is

encoded in the OFOS-5 primer. Jun is PCR amplified using pOK1 as template and oligonucleotides OJUN-5 and OJUN-3 as primers.

OFOS-5 5'- GGG*GATATC*CACCACCACCACCACCACCTGCGGTGGTCTGACC

OFOS-3 5'- GGG*GAATTC*CAACCACCGTGTGCCG

OJUN-5 5'- GGG*GATATC*GGTGGTCGGATCGCC

OJUN-3 5'- GGG*GAATTC*ACCACCGTGGTTCATGAC

The hot-start procedure is used. A step-wise touch-down PCR is applied: 92°C, 1 min; 58-52°C, ΔT = 2°C, 1 min; 72°C, 1 min. This is followed by 26 cycles (92°C, 1 min; 52°C, 1 min; 72°C, 1 min).

The PCR products are purified using QIAquick kit (Qiagen) and eluted in ddH_2O. They are then overnight digested with *Eco*RI and *Eco*RV.

The p75 fragment is also PCR amplified using pUC18-IMPp75 (see Example 4) as template and oligonucleotides OP75-5 (where His_6 is encoded) and OP75-3 as primers:

OP75-5 5'- GGG*GATATC*CACCACCACCACCACCACAAGAGGTGGAACAGC

OP75-3 5'- GGG*GAATTC*CACTGGGGATGTGGCAG

The same PCR and restriction digestion conditions as above are applied.

The IL-16 fragment is amplified from the cDNA clone **pcDNA3-ILHu1** (M. Baier, Paul Ehrlich Institute, Germany; Baier *et al.*, 1995; Bannert *et al.*, 1996), using OIL16-5 (where His_6 is encoded) and OIL16-3 as primers.

OIL16-5 5'- GGG*GATATC*CACCACCACCACCACCACCCCGACCTCAACTCCTC

OIL16-3 5'- GGG*GAATTC*GGAGTCTCCAGCAGCTG

The same PCR and restriction digestion conditions as above are applied.

In all cases, the fragments are readily amplified and digested.

2.5: Cloning into intermediate vectors

The digested PCR fragments are gel-purified (QIAquick kit, Qiagen) and eluted into TE buffer. The *Eco*RV/*Eco*RI fragment of pIG1 vector (Ge *et al.*, 1995) is also isolated. The digested PCR fragments of fos, p75, and IL-16 are ligated into the vector fragment, and the ligated vectors transformed into TG1 cells.

The constructs in the pIG1 vector contains the OmpA signal sequence fused in-frame with the constructs.

The correct clones are screened and confirmed by sequencing. They are then *Xba*I/*Hin*dIII digested, and the fragments are isolated.

2.6: Cloning into the expression vectors

The isolated fragments from 2.3 are inserted into pBS13 also excised with *Xba*I/*Hin*dIII, resulting in vectors **pYING1-C1** (Fos), **pYING1-C2** (p75), **pYING1-C3** (IL-16) (see Figure 5). The fragment containing jun is cloned into pIG10.3 vector *via* *Eco*RV/*Eco*RI resulting in **pYANG3-A** (see Figure 6). The anti-p75 peptides pe2, pe3 and pe10 (see Example 4) are cloned into pIG10.3 via *Xba*I/*Hin*dIII, resulting in vectors **pYANG3-Ape2**, **-Ape3** and **-Ape10**, respectively (see Figure 6).

2.7: Selection of correct pairing *via* His-tag

TG1 cells are transformed with the combination of pYANG3-A + pYING1-C1, or pYANG3-A + pYING1-C2, or pYANG3-A + pYING1-C3, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C1, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C2, or (pYANG3-Ape2, -Ape3 and -Ape10) + pYING1-C3, thus creating all possible combinations separately to ensure the presence of each of them in the selection experiment. The transformed cells are plated on ampicillin/chloramphenicol-containing LB agar plates, and colonies with double resistance (Ap^R/Cm^R) are selected.

The colonies are scraped off the plates and used to inoculate 2xYT medium (Amp/Cm) and shaken at 37°C for 3 hrs. The cultures are induced (1 mM IPTG) at 30°C for 1 hr and infected with R408 (Stratagene) at 37°C for 30 min. The cultures are shaken at RT for 3 hrs, kanamycin is added and shaking continued at RT overnight.

The phage particles are harvested from the overnight cultures, mixed and PEG-precipitated. The phages are directly selected on immobilized Ni-NTA (NI-NTA HisSorb Strips, Qiagen). The eluted phages are used to infect TG1 cells, which are plated on ampicillin/chloramphenicol-containing LB agar plates, and colonies with double resistance (Ap^R/Cm^R) are selected.

The phagemids of selected clones are isolated and analyzed by restriction digest (see Figure 7.a) and used as templates for PCR screening. Primer OPEP5L is used to amplify the pYANG3-Ape2, -Ape3 and -Ape10 constructs specifically (see Figure 7.b).

OPEP5L 5'- GACTACAAAGATGTCGACTG

There is a specific enrichment of constructs of correct pairing (Table 2).

Example 3: Interactive screening of *E. coli* genomic DNA libraries (Polyphage/tag system)

3.1: Principle (see Figure 2)

Instead of using two model libraries as in Example 2, a genomic DNA library of *E. coli* is prepared to be screened against itself to identify interacting *E. coli* peptides or proteins.

3.2: Construction of display and expression vectors for genomic DNA

Expression vectors are constructed having a blunt-end restriction site SmaI inserted either in front of His-tag, Strep-tag (Schmidt and Skerra, 1994) or the C-terminal domain of gIII (gIIIc) *via* oligonucleotide cassettes or PCR.

The self-complementary oligonucleotides OHIS5 & OHIS3, and OSTREP5 & OSTREP3, are used to create ds DNA cassettes encoding the His-tag, and the Strep-tag, respectively.

OHIS5 5'- AATTCCCCGGGCACCACCACCACCACCACTGATA

OHIS3 5'- AGCTTATCAGTGGTGGTGGTGGTGGTGCCCGGGG

OSTREP5 5'- AATTCCCCGGGTCTGCTTGGCGTCACCCGCAGTTCGGTGGT-TGATA

OSTREP3 5'- AGCTTATCAACCACCGAACTGCGGGTGACGCCAAGCAGACC-CGGGG

The cassettes upon phosphorylation and annealing recreate the *Eco*RI and *Hin*dIII sites. The cassettes are inserted into pIG1 and pIG3 vectors (Ge *et al.*, 1995) cut by the same restriction enzymes. The resulting vectors are pING1-A1, pING3-A1 (for His tag in pIG1 and pIG3 vectors) and pING1-A2, pING3-A2 (for Strep-tag), respectively. The correct vectors are screened for the presence of *Xma*I site (isoschizomer of *Sma*I) and the constructs are confirmed by sequencing. The *Xba*I/*Hin*dIII fragments of these vectors are inserted into pBS13 vector, linearized with the same enzymes, resulting in vectors **pING1-C1**, **pING3-C1** and **pING1-C2**, **pING3-C2,** respectively (see Figure 8).

The gIIIc fragment containing the SmaI site is generated from PCR amplification of pIG10.3 vector using primers OGIII5 and OGIII3, where OGIII3 anneals 3' of the gene III in the vector:

OGIII5 5'- CG*GAATTC*CCCGGGGAGCAGAAGCTGATC

OGIII3 5'- TTTTTCACTTCACAGGTC

Three rounds of PCR are performed with a hot-start: 92°C, 1 min; 46°C, 1 min; 72°C, 1.5 min. This is followed by 30 rounds of: 92°C, 1 min; 50°C, 1 min; 72°C, 1.5 min.

The PCR product is purified (QIAquick) and digested with *Eco*RI and *Hin*dIII. The fragment is gel-purified (QIAquick) and ligated into pIG10.3. The sequence of the resulting vector, **pONG3-A** (see Figure 8), is confirmed by restriction analysis and by sequencing.

3.2: Selection of Interacting Pairs from *E. coli* Genomic DNA via His-tag

Genomic DNA of *E. coli* strain XL-1 Blue (Stratagene) is isolated using the Blood & Cell Culture DNA Maxi kit (Qiagen) and eluted in TE buffer (pH 8.0). 200 µg of the DNA is

taken and sonicated (50 cycles, 270 mA, 0.5 s/stroke). The fragmented DNA (average size: max. 0.7 kB) is blunt-ended by a fill-in reaction with T4 DNA polymerase.

Vectors pING1-C1 and pONG3-A are digested with *Eco*RV and *Sma*I, the vector fragments are gel-purified (Qiagen). The vector fragments are then ligated with the blunt-ended genomic DNA at 16°C overnight. The ligation mixtures are taken to transform TG1 cells.

The pING1-C1 and pONG3-A transformants are scratched from the plate and used to inoculate 2xYT medium containing Cm/glucose or Amp/glucose, respectively. The pING1-C1 culture is infected with helper-phage (VCSM13 or M13k07) and phage particles are isolated. These phage particles are used to infect log-phase cells containing the pONG3-A library. The resulting culture is plated out on large Amp/Cm/glucose plates.

The colonies are scratched from the surface of the plates above and transferred to 2xYT medium containing Amp/Cm. After 30 min shaking at 37°C, the culture is then induced (1 mM IPTG) for 30 min, infected with helper-phage at 37°C for 30 min and shaken at RT overnight.

The phage particles are harvested from the overnight culture and PEG-precipitated. They are selected on immobilized Ni-NTA (NI-NTA HisSorb Strips, Qiagen). The eluted phages are used to infect log-phase TG1 cells. Selected protein pairs are characterised by determination of their corresponding DNA sequences.

Example 4: Polyphages and Selection of Correct Pairing Interactions *via* SIP

4.1: Principle (see Figure 10)

The purpose of this experiment is to show that from a combination of 2 libraries one can isolate and identify the correct interacting pairs using the SIP (Selectively Infective Phage: Krebber *et al.*, 1995; the term "IMP" used in the experimental section denotes "Infectivity mediating particle" comprising the N-terminal domains of the gene III protein

of filamentous phage) selection system, and recover the information about both interacting partners *via* the formation and selection of polyphage particles. The library members forming interacting pairs with members of the corresponding library are being 'doped' with library members that do not interact with members of the corresponding library, and thus should not give a positive SIP selection.

4.2: Construction of vectors

4.2.1: fhag1A (see Figure 11)

a. The phage vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRV and XmnI. The 1.1 kb fragment containing the anti-HAG Ab gene is isolated by agarose gel electrophoresis and purified with a Qiagen gel extraction kit. This fragment is ligated into a pre-digested pIG10.3 vector (EcoRV-XmnI). Ligated DNA is transformed into DH5a cells and positive clones are verified by restriction analysis. The recombinant clone is called **pIGhag1A**. All cloning described above and subsequently are according to standard protocols (Sambrook *et al.*, 1989)

b. The vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRV and StuI. The 7.9 kb fragment is isolated and self-ligated to form the vector **fhag2.**

c. The chloramphenicol resistance gene (CAT) assembled *via* assembly PCR (Ge and Rudolph, 1997) using the the template pACYC (Cardoso and Schwarz, 1992) (Figure 11a shows the functional map and the sequence of the CAT gene) is amplified by the polymerase chain reaction (PCR) with the primers:

CAT_BspEI(for): 5' GAATGCTCATCCGGAGTTC

CAT_Bsu36I(rev): 5' TTTCACTGGCCTCAGGCTAGCACCAGGCGTTTAAG

d. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with BspEI and Bsu36I then ligated into pre-digested fhag2 vector (BspEI-Bsu36I; 7.2 kb fragment) to form **fhag2C**.

e. The vector fhag2C is digested with EcoRI and the ends made blunt by filling-in with Klenow fragment. The flushed vector is self-ligated to form vector **fhag2CdelEcoRI**.

f. pIGHAG1A is digested with XbaI and HindIII. The 1.3 kb fragment containing the anti-HAG gene fused with the C-terminal domain of filamentous phage pIII protein is isolated and ligated with a pre-digested fhag2CdelEcoRI phage vector (XbaI-HindIII; 6.4 kb) to create the vector **fhag1A**

4.2.2: fjun1A (see Figure 12)

a. The EcoRV site of pIG10.3 is converted to a SalI site by oligonucleotide site-directed mutagenesis (Sambrook *et al.*, 1989) with primer:
Sall9-9primer(rev) 5'CTGAATGTCGACATCTTTGTAGTC3'
The mutated pIG10.3 is called **pIG10.3 SalI.**

b. The jun leucine-zipper domain from **pOK1** (Grammatikoff *et al.*, 1994) is amplified by PCR with the primers:
jun2(for): 5'ACGCGTCGACGCCGGTGGTCGGATCGCCCGG3'
jun2(rev): 5'AATTCGGCACCACCGTGGTTCATGACT3'

c. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with SalI and EcoRI, then ligated into pre-digested pIG10.3SalI vector (SalI-EcoRI) to form the vector **jun-pIG10.3SalI.**

d. The vector jun-pIG10.3SalI is digested with XbaI and EcoRI. The 0.14 kb fragment is ligated into the pre-digested vector fhag1A (XbaI-EcoRI; 7kb) to form the vector **fjun1A**.

4.2.3: fjun1B (see Figure 13)

a. The DNA encoding the C-terminal domain including the long linker separating it from the amino terminal domain of the filamentous phage pIII (gIII short) is amplified by PCR using pOK1 (Grammatikoff *et al.*, 1994) as template with the primers:
gIII short(for): 5'GCTTCCGGAGAATTCAATGCTGGCGGCGGCTCT3'
gIII short(rev): 5'CCCCCCCAAGCTTATCAAGACTCCTTATTACG3'

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with EcoRI and HindIII, then ligated into pre-digested fhag1A vector (EcoRI-HindIII) to form the vector **fjun1B.**

4.2.4: fpep2_1b, fpep3_1B, fpep10_1b (see Figure 14)

a. These constructs are obtained from a peptide library screened against the intracellular domain of p75, the low affinity receptor of NGF, in a SIP experiment.

b. A peptide library cassette of cyclic peptides with length variants of 6-16 amino acids is prepared from the oligos:

Groprim: 5'-CATGAATTCGGATCCTCC-3'

Gron10: 5'-CTATGGCGCGCCTGTCGACTGT(M)$_{6\text{-}16}$TGTGGTGGTGGAGGATC-CGAATTCATG-3'

where M is a mixture of 19 trinucleotide codons (Virnekäs *et al.*, 1994), excluding the one coding for Cys. The length variation is achieved by coupling 6 trinucleotide positions using the standard coupling procedure, and, for the next 10 coupling cycles, by omitting the capping step during DNA synthesis and by diluting the trinucleotide mixture to achieve stepwise coupling yields of 50%.

The oligos are annealed and filled in with the Klenow fragment of DNA polymerase I to form a double-stranded DNA cassette with standard methods (Sambrook *et al.*, 1989). The cassette is digested with SalI-EcoRI, purified with Qiaex DNA gel extraction kit, and ligated to pre-digested fjun1B vector (SalI-EcoRI) to form the peptide library. The ligated peptide library is transformed into competent DH5a cells harboring pUC18/IMP-p75 (see below) and plated on Luria Broth (LB) (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

c. The Amp^r Cm^r colonies are scraped with LB, and 1 ml of suspension is used to inoculate 25 ml LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin + 1 mM IPTG). The culture is incubated overnight at room temperature.

d. The supernatant is separated from the cells by centrifugation (10,000 RPM, 10 min., 4°C). 5 ml of 30% PEG/3M NaCl are added to the supernatant and mixed 100 times. After 1 hour on ice, the phage precipitate is collected by centrifugation (10,000 RPM, 10 min., 4C). The pellet is resuspended in 1 ml TBS buffer. The suspension is filtered with a 0.45 micron filter (Sartorius).

e. 100 µl of log phase K91 cells (or any male E. coli cells (F-pilus containing)) are infected with 10 µl of phage supernatant, plated on LB (30 µg/ml chloramphenicol) and incubated overnight at ambient temperature.

f. Chloramphenicol-resistant transductants are picked, and overnight cultures are prepared to isolate DNA for sequencing. From the sequencing, fpep2_1b, fpep3_1B, fpep10_1b containing peptides pe2, pe3, and pe10 are identified.

pe2: 5'-TGTTTTTTTCGTGGTGGTTTTTTTAATCATAATCCTCGTTATTGT-3'

(CysPhePheArgGlyGlyPhePheAsnHisAsnProArgTyrCys)

pe3: 5'-TGTATTGTTTATCATGCTCATTATCTTGTTGCTAAGTGT-3'

(CysIleValTyrHisAlaHisTyrLeuValAlaLysCys)

pe10: 5'-TGTTCTTATCATCGTCTTTCTACTCGTGTTTGT-3'

(CysSerTyrHisArgLeuSerThrArgValCys)

4.2.5: fNGF1B (see Figure 15)

a. The DNA encoding the nerve growth factor (NGFI) gene is amplified from pXM NGF (Ibanez *et al.*, 1992) as template with the primers:

NGF(for): 5'AAAAAAGTCGACTCATCCACCCACCCAGTC3'

NGF(rev): 5'AGGAATTCGCCTCTTCTTGCAGCCTT3'

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with SalI and EcoRI, then ligated into pre-digested fjun1B vector (SalI-EcoRI) to form the vector **fNGF1B.**

4.2.6: pUC19/IMP-HAG (see Figure 16)

a. The vector f17/9-hag (Krebber *et al.*, 1995) is digested with EcoRI and HindIII. The 1.4 kb fragment containing the gene fusion of the IMP with the HAG peptide, is isolated and cloned into pre-digested **pUC19** (EcoRI-HindIII) to form the vector **pUC19/IMP-HAG**

4.2.7: pUC18/IMP-p75 (see Figure 17)

a. The intracellular domain of p75 containing the C-terminal 142 amino acids is amplified from the cDNA clone of p75 (Chao *et al.*, 1986) as template with the primers:

p75(for): 5' GCTGGCCCGTACGACAAGAGGTGGAACAGCTGC

p75(rev): 5' TCTCGAAGCTTATCACACTGGGGATGTGGC

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with BsiWI and HindIII, then ligated into pre-digested pUC19 vector (BsiWI-HindIII) to form the vector **pUC19/IMP-p75.**

c. The vector pUC19/IMP-p75 is digested with XbaI and HindIII. The 1 kb fragment is isolated and cloned into the pre-digested **pUC18** vector (XbaI-HindIII) to form the vector **pUC18/IMP-p75.**

4.2.8: pUC18/IMP-IL16 (see Figure 18)

a. The IL16 gene is amplified from the clone **pcDNA3-ILHu1** (M. Baier, Paul Ehrlich Institute, Germany; Baier *et al.*, 1995; Bannert *et al.*, 1996) as template with the primers:

f1Bsu36Ifor: 5'AGACTGCCTCAGGCCAGCCCGACCTCAACTCC3'

f3HindIIIrev2: 5'ATATATAAGCTTTTAGGAGTCTCCAGCAGC3'

b. The PCR is done following standard protocols (Sambrook *et al.*, 1989). The amplified product is digested with Bsu36I and HindIII, then ligated into pre-digested pUC18/IMP-p75 vector (Bsu36I-HindIII) to form the vector **pUC18/IMP-IL16.**

4.3: In vivo SIP with co-transformation and polyphage

4.3.1: Combining 2 libraries (Library 1 is fused with gIII while Library 2 is fused to the IMP).

10 ng each of fjun1B, fjun1A, fpep3_1B, fhag1A, fNGF1B with 500 ng each of pUC18/IMP-p75, pUC18/IMP-HAG, pUC18/IMP-IL16 are co-transformed into DH5a cells by electroporation. The cells are plated on Luria Broth (LB) (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

The Amp^r Cm^r colonies are scraped with LB and 1 ml of suspension is used to inoculate 25 ml LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin + 1 mM IPTG) followed by incubation overnight at room temperature.

4.3.2: In vivo SIP. The supernatant from the cells is separated by centrifugation (10,000 RPM, 10 min., 4°C). 5 ml of 30% PEG/3M NaCl are added to the supernatant and mixed 100 times. After 1 hour on ice, the phage precipitate is collected by centrifugation (10,000 RPM, 10 min., 4°C). The pellet is resuspended in 1 ml TBS buffer, and the suspension is filtered through a 0.45 micron filter (Sartorius).

200 μl of phage supernatant are used to infect 1.8ml of log phase K91 cells (or any male E. coli cells (F-pilus containing)), and the cells are plated on LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) and incubated overnight at ambient temperature.

4.3.3: Testing of infectious polyphage DNA patterns and infectity. Twenty individual Amp^r Cm^r colonies are used to inoculate 5 ml LB (30 μg/ml chloramphenicol + 100 μg/ml ampicillin) in each case and incubated at ambient temperature overnight. Plasmid and RF DNA are isolated from each clone with a Qiagen Miniprep DNA kit. Clones are analysed by restriction analysis with restriction enzymes XbaI and HindIII together with appropriate buffers as supplied and instructed by the manufacturer. The restriction

digests are run in a 0.8% TBE agarose gel at constant voltage of 100V for 1.5 hours. The restriction patterns, together with the relative intensity of the bands (because the phage vectors (fjun1B, fjun1A, fpep3_1B, fNGF1B, fhag1A) have significantly lower copy numbers than the plasmid vectors) allow to identify correctly interacting pairs. For the pair fhag1A+pUC19/IMP-HAG, an XbaI-HindIII digest will yield a 6.5 kb, 3.3 kb, 1.3 kb, and 0.7 kb fragments, while for the pair fpep3_1B+pUC18/IMP-p75, the same digest will yield 6.3 kb, 2.8 kb, 1kb, and 0.7kb fragments. A problem though is to distinguish the potential non-cognate combinations of fjun1B or fjun1A with pUC18/IMP-p75 because they would give similar patterns as the fpep3_1B+pUC18/IMP-p75. To further resolve this, the clones containing identical patterns can be re-digested with BamHI-HindIII. The fjun1A or fjun1B in combination with pUC18/IMP-p75 would yield only 4 fragments - 4.1 kb and 2.9 kb , 2.6 kb , 1.2 kb fragments - while the cognate pair fpep3_1B+pUC18/IMP-p75 will yield 5 fragments - 3.5 kb, 2.9 kb, 2.6 kb, 1.2 kb, 0.5 kb.

To further prove that cognate interacting pairs have been selected, the ability of the clones to form selectively-infective phage particles is tested. Only clones with a cognate pair can form infectious phages. The supernatant from the overnight culture of the individual clones is filtered with a 0.45 micron filter (Sartorius). Ten microliters of phage supernatant are mixed with 100 µl of log phase K91 cells (or any male E. coli cells (F-pilus containing)) for 10 minutes at 37°C. The suspension is plated on LB (30 µg/ml chloramphenicol) and incubated overnight at 37°C. The result is shown in Table 3.b.

In summary (see Figure 19), the results from the above example indicate that among 19 clones analyzed, 8/19 have the cognate pair fpep3_1B+pUC18/IMP-p75 and produce selectively-infective phage; 1/19 has the fhag1A+pUC19/IMP-HAG combination and produces selectively-infective phage.

Example 5: Combination of Multiple Libraries into a Single Phagemid Vector through Recombination, Screening *via* tag system

5.1: Principle (see Figure 20)

To be able to retrieve the genetic information for cognate protein pairs selected *via* a tag fused to one of the partners, two separate libraries in phagemid vectors are constructed containing the *lox* recombination promoting sites and recombined on one phagemid by action of the *cre* recombinase in an *in vivo* recombination.

5.2: Vector construction

Both *lox*P and *lox*P511sites (Hoess *et al.*, 1986) are inserted in tandem into the region flanked by the ColE1 ori and β-lactamase in vector pING1-C1, whereas in vector pONG3-A, the *lox*P site is cloned upstream of the *Xba*I site and the *lox*P511 downstream of the *Hind*III site. Therefore, the genomic DNAs to be cloned are flanked by the *lox*P and *lox*P511 sites.

5.3: Library construction and recombination

The libraries are prepared as in Example 3. The phagemids in the double-resistant clones are recombined through the *cre* recombinase which either is encoded in the phagemid being inducible (Tsurushita *et al.*, 1996), or is transferred through P1 phage infection (Rosner, 1972; Waterhouse *et al.*, 1993). Phages are prepared from the recombined clones by helper phage infection and used to infect new *E. coli* cells (*cre*$^-$).

5.4: Selection

The phage particles are prepared from the Cm^R clones and subjected to His-tag selection as in Examples 2 and 3. The sequences encoded in each phagemid, which now contains members of both libraries, can be determined by sequencing using primers specific for myc-tag region (library 1) and His-tag region (library 2).

Example 6: SIP-based library vs. library screening via *in vitro* recombination of separately constructed libraries into one phage vector

6.1: Principle (see Figur 21)

To be able to retrieve the genetic information for cognate protein pairs selected by SIP interaction *in vivo*, two separate libraries in phage and plasmid vectors are constructed and recombined by co-ligation in an *in vitro* recombination.

6.2: Construction of Libraries A and B

Library A encodes 2 members, namely a single chain Fv antibody against a peptide derived from hemagglutinin (fαhag) and the leucine zipper domain derived from the jun transcription factor (fjun), both N-terminally fused to the C-terminal domain of gIII from filamentous phage and preceded by the ompA signal sequence followed by the Flag epitope.

Library B encodes 3 members on plasmid vectors of the pUC series, namely the hemagglutinin peptide to which the above αhag antibody binds (pUC19-IMPhag), the leucine zipper domain of the fos transcription factor (pUC18-IMPfos) which heterodimerizes with jun via this domain, and the intracellular domain of the low affinity nerve growth factor receptor (pUC18-IMPp75), as a negative control which does not interact with library A members, all fused to the infectivity-mediating N-terminal domains of phage gIII protein, preceded by the gIII signal sequence.

Library A members are cloned into a fd phage vector which also contains downstream of the library A insertion site the N-terminal domains (N1-N2) of gIII, followed by the cloning sites *Bsi*WI and *Hin*dIII to allow in-frame insertion of library B members.
Library A construct fαhag is identical to the f17/9-hag fd phage vector (Krebber *et al.*, 1995) and serves as basis for construction of fjun. The jun leucine zipper together with amino acids 290 to 326 of the C-terminal part of gIII is PCR-amplified (primers FR620 and FR621, containing *Eco*RV and *Sfi*I sites, respectively) from the construct fjun1B (containing the jun leucine zipper fused to amino acids 290 to 493 of gIII) generated in Example 4. The resulting PCR fragment is ligated directionally into EcoRV/SfiI-digested

f17/9-hag vector in frame with amino acids 327 to 493 of the gIII C-terminal domain resulting in vector **fjunhag** (see Figure 22).

Generation of library B constructs pUC19-IMPhag and pUC18-IMPp75 is described in Example 4. To construct pUC18-IMPfos, amino acids 219 to 272 of the N-terminal part of gIII together with the fos leucine zipper are PCR-amplified (primers FR618 and FR619, containing BsiWI and HindIII sites, respectively) from the pOK1 phagemid vector (Grammatikoff *et al.*, 1994). The resulting PCR fragment is ligated directionally into BsiWI/HindIII-digested pUC18-IMPp75 to create **pUC18-IMPfos** (see Figure 17).

Primers:

FR618: 5'CGCCGTACGGCGGCTCTGGTGGTGGTTCTGGTGGC3'

FR619: 5'CCCAAGCTTTTAGACTAGCTGACTAGAAGATCTGC3'

FR620: 5'CGCGATATCGTCGACGCCGGTGGTCGGATCGCC3'

FR621: 5'CGCGGCCCCCGAGGCCCCACCACCGGAACCGCCTCCC3'

6.3: Preparation and recombination of library A and B and selection of interacting protein pairs by SIP

Non-covalent, cognate interactions of αhag antibody with hag peptide (Krebber *et al.* 1995) and of fos and jun leucine zipper domains (Grammatikoff *et al.*, 1994) generates infective SIP phage. Thus, from the six possible combinations of members of the model libraries A and B (**fαhag-hag**, fαhag-fos, fαhag-p75, **fjun-fos**, fjun-hag, fjun-p75), only two combinations (cognate pairs in bold) should be selected by *in vivo* SIP. To recombine the library members in all possible permutations, library A is linearized by digestion with BsiWI/HindIII to prepare it for random incorporation of library B members, prepared by mass-excision with BsiWI/HindIII from the construct B pool described above. After co-ligation of the mass-excised library B fragments into library A vectors, the sample is transformed into competent E.coli cells, plated onto chloramphenicol-containing LB agar plates and grown overnight at 37°C. The recombined library size can be determined by plating serial dilutions of the transformation and can be compared to

the complexities of the individual libraries A and B. The total recombined library is scraped from the plates in LB medium and used to inoculate an appropriate volume of chloramphenicol-selective LB-medium supplemented with 1 mM IPTG. After growth at 30°C overnight with constant shaking to allow production of SIP phages, the bacteria are pelleted by centrifugation and phages present in the supernatant are precipitated on ice for one hour by addition of 0.25 volumes of 20% PEG/2.5 M NaCl. The phages are pelleted by centrifugation for 30 min at 10 000 x g and 4°C. The pellet is resuspended in an appropriate volume of 1 x TBS buffer and filtered through a 0.45 µM filter. Serial dilutions of this filtrate are used to infect F^+ E.coli cells. The double-stranded, replicative form phage DNA is prepared from resulting transductant colonies by standard methods and analyzed by restriction digest and sequencing for the presence and identity of library A and B members. Furthermore, the supernatant of transductant colonies is analyzed for the presence of infective SIP phages to confirm that protein-protein interaction of a particular pair selected from the recombined libraries A and B is responsible for SIP phage infectivity.

Alternatively, the model libraries A (2 members) and B (3 members) are used to construct all possible combinations (listed above) individually, and equal amounts (50 ng) of each of the 6 combinations can be co-transformed into competent E. coli cells followed by the steps listed above. The distribution of individual constructs after co-transformation as well as the distribution of transductants resulting from the model library can be analyzed as described above. The selective recovery of phage constructs which co-encode cognate protein pairs demonstrates the feasibility of SIP-based selection of binding partners after an appropriate recombination event.

Example 7: 'Spatial' *in vivo* SIP

7.1: Principle (see Figure 23)

Coupling of information about members of interacting peptides or proteins is achieved by having a spatial relationship between the particles displaying the selectable or

screenable property (in this example phages for the SIP experiment) and the package containing the genetic information for the individual library members (in this example the *E. coli* cell secreting the phage particle being screened), i. e. a correlation between the phage being examined and the position of the corresponding *E. coli* host on the master plate.

7.2: Combining 2 libraries (Library A is fused with gIII while library B is fused to the IMP)

10 ng each of fjun1B, fjun1A, fpep3_1B, fhag1A, fNGF1B are co-transformed with 500 ng each of pUC18/IMP-p75, pUC19/IMP-HAG, pUC18/IMP-IL16 into DH5a cells by electroporation. The transformants are plated on LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) and incubated overnight at ambient temperature.

7.3: Screening of co-transformants by SIP

From the master plate of co-transformants, each of the co-transformants are labelled and inoculated separately into 5 ml LB (30 µg/ml chloramphenicol + 100 µg/ml ampicillin) and incubated overnight at ambient temperature.

Plasmid and RF DNA are isolated from each clones with a Qiagen Miniprep DNA kit. Clones are analysed by restriction analysis with restriction enzymes XbaI and HindIII together with appropriate buffers as supplied and instructed by the manufacturer. The restriction digests are run in a 0.8% TBE agarose gel at constant voltage of 100 V for 1 to 2 hours. Restriction patterns allow discrimination of the particular clones.

The supernatant from the overnight culture of the individual clones is filtered with a 0.45 micron filter (Sartorius). Ten microliters of phage supernatant are mixed with 100 µl of log phase K91 cells (or any male E. coli cells (F-pilus containing)) for 10 minutes at 37°C. The suspension is plated on LB (30ug/ml chloramphenicol) and incubated overnight at 37°C.

A positive co-transformant (i.e. contains the correct interacting pair) has a corresponding correct restriction pattern and is capable of producing infectious phages, that are incapable of secondary or subsequent infections. Polyphage particles being capable of such infections, and containing the genetic information of an interacting pair as well, can readily be identified by their restriction digest pattern.

Example 8: *E. coli* display

8.1: Principle (see Figure 24)

Two libraries are introduced into *E.coli* cells, with expressed members of library A (such as antibody, peptide, or cDNA libraries) being presented at the surface of the cells. In those cases where interacting pairs are formed, members of library B (such as antibody, peptide, or cDNA libraries) are transported in the complex with its cognate partner to the surface of the cell as well, thus displaying a selectable or screenable property such as a tag. Selected cells contain the information for both interacting partners.

8.2: Preparation of Library A

A thioredoxin peptide library is prepared as fusions to the *E. coli* flagellin in the pFLITRX vector essentially as described (Lu *et al.*, 1995).

8.3: Preparation of Library B

An cyclic, variable-length peptide library including a FLAG epitope (Hopp *et al.*, 1988; Knappik and Plückthun, 1994) is prepared essentially as described in Example 4.2.4, and cloned in the pTERM vector, a modified version of the pto2H10a3s vector (Krebber *et al.*, 1996) containing a chloramphenicol-resistance gene instead of the kanamycin-resistance gene. The pTERM vector can be assembled by standard methods starting from pto2H10a3s. This cyclic peptide library is packaged by infection with a helper phage (M13K07 or VCSM13) by standard methods (Sambrook *et al.*, 1989).

8.4: Combination of Library A and Library B

An aliquot of the *E. coli* cells containing Library A is used to inoculate 50 ml LB (100 µg/ml ampicillin) and incubated at ambient temperature until the OD600 reached 0.4. The cells are infected with phages containing Library B at a multiplicity of infection (MOI) of 10. After 30 min of infection, the cells are collected by centrifugation (5000 RPM, 10 minutes, 4°C) and resuspended in 1 ml LB. The suspension is plated on M9 media (+ 1 mM $MgCl_2$, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol).

8.5: Selection of interacting pairs

The Amp^r Cm^r colonies are scraped with M9 media (+ 1 mM $MgCl_2$, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol), and an aliquot of the suspension is used to inoculate 25 ml M9 media (+ 1 mM $MgCl_2$, supplemented with 0.5% glucose, 0.2% casamino acids, 100 µg/ml ampicillin, 30 µg/ml chloramphenicol) and incubated at 37°C until saturation. Selection is performed essentially as described (Lu *et al.*, 1995), the modification being that the antibody used for selection is the M1 anti-FLAG antibody (Kodak).

Individual enriched Amp^r Cm^r colonies are isolated and the sequences of the corresponding interacting peptide(s) and cyclic peptide(s) are determined by DNA sequencing. To confirm that the encoded peptide and cyclic peptide form a cognate pair, each of the clones is tested for enrichment based on the selection method described above, whereby the Amp^r Cm^r colonies bind to the M1 anti-FLAG antibody in a single round of selection.

Literature:

Baier *et al.*, 1995, *Nature* 378, 563

Bannert *et al.*, 1996, *Nature* 381, 30

Cardoso and Schwarz, *J. Appl. Bacteriol.* 72 (1992) 289-293

Crameri and Suter, 1993, *Gene* 137, 69-75

Ge *et al.*, 1995, *Antibody Engineering*, 2nd ed., 229-266

Ge and Rudolph, *BioTechniques* 22 (1997) 28-29

Gramatikoff *et al.*, *Nucleic Acids Res.* 22 (1994) 5761-5762

Hochuli *et al.*, 1988, *Bio/Technology* 6, 1321-1325

Hoess *et al.* 1986, *Nucleic Acids Res.* 14, 2287-2300

Hopp *et al.*, 1988, *Bio/Technology* 6, 1204-1210

Ibanez *et al.*, 1992, *Cell* 69, 329-341

Knappik and Plückthun, 1994, *BioTechniques* 17, 754-761

Krebber *et al.*, 1995, *FEBS Letters* 377, 227-231

Krebber *et al.*, 1996, *Gene* 178, 71-74

Lindner *et al.*, 1992, *Methods: A companion to Methods Enzymol.* 4, 41-56

Lu *et al.*, 1995, *Bio/Technology* 13, 366-372

Rosner, 1972, *Virology* 48, 679-689

Sambrook *et al.*, 1989, *Molecular Cloning: a Laboratory Manual*, 2nd ed.

Schmidt and Skerra, 1994, *J. Chromatogr. A* 676, 337-345

Tsurushita *et al.*, 1996, *Gene* 172, 59-63

Virnekäs *et al.*, 1994, *Nucleic Acids Res.* 22, 5600-5607

Waterhouse *et al.*, 1993, *Nucleic Acids Res.* 21, 2265-2266

CLAIMS

1. A method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

   (a) providing a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

   (b) providing a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (a), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (a) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (a) and (b), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

   (c) optionally, providing additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (a) and/or step (b), wherein the vector molecules employed for the

production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (a) and (b) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (c) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (a) and/or (b), upon the interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

(d) expressing members of said libraries of recombinant vectors or nucleic acid sequences mentioned in steps (a), (b) and optionally (c), in appropriate host cells so that at least one interaction is established;

(e) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;

(f) optionally, carrying out further screening, selection and/or purification steps; and

(g) identifying said nucleic acid sequences encoding said (poly)peptides.

2. A method for identifying a plurality of nucleic acid sequences, said nucleic acid sequences each encoding a (poly)peptide capable of interacting with at least one further (poly)peptide encoded by a different member of said plurality of nucleic acid sequences, comprising the steps of:

(a) expressing in appropriate host cells

(aa) nucleic acid sequences contained in a first library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides;

(ab) nucleic acid sequences contained in a second library of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with further (poly)peptides as mentioned in step (aa), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in step (aa) and wherein at least one of said properties displayed by each of said vector molecules and/or the recombinant inserts used in steps (aa) and (ab), upon the interaction of a (poly)peptide from said first library with a (poly)peptide from said second library together generate a screenable or selectable property;

(ac) optionally, nucleic acid sequences contained in additional libraries of recombinant vector molecules containing genetically diverse nucleic acid sequences comprising a variety of nucleic acid sequences encoding (poly)peptides capable of interacting with or causing interaction of (a) further (poly)peptide(s) as mentioned in step (aa) and/or step (ab), wherein the vector molecules employed for the production of said recombinant vector molecules and/or the recombinant inserts display properties that are phenotypically distinguishable from those of the vector molecules and/or the recombinant inserts used in steps (aa) and (ab) and, optionally, at least one of said properties displayed by said vector molecule and/or the recombinant inserts used in step (ac) together with at least one of said properties displayed by either said vector molecule and/or said recombinant inserts used in steps (aa) and/or (ab), upon the

interaction of a (poly)peptide from said additional library with either a (poly)peptide from said first library and/or a (poly)peptide from said second library generate a screenable or selectable property;

so that at least one interaction is established;

(b) selecting for the generation of said screenable or selectable property representing the interaction of said (poly)peptides;
(c) optionally, carrying out further screening, selection and/or purification steps; and
(d) identifying said nucleic acid sequences encoding said (poly)peptides.

3. The method according to claim 1 or 2, wherein said screenable or selectable property is expressed extracellularly.

4. The method according to any one of claims 1 to 3 wherein said recombinant vector molecules in step (a)/(aa) give rise to a replicable genetic package (RGP) displaying said (poly)peptides at its surface.

5. The method according to claim 4, wherein said recombinant vector molecule is a recombinant phage, phagemid or virus.

6. The method according to claim 5, wherein said phage is
(a) one of the class I phage fd, M13, If, Ike, ZJ/2, Ff;
(b) one of the class II phage Xf, Pf1, and Pf3;
(c) one of the lambdoid phages, lambda, 434, P1;
(d) one of the class of enveloped phages, PRD1; or
(e) one of the class paramyxo-viruses, orthomyxo-viruses, baculo-viruses, retro-viruses, reo-viruses and alpha-viruses.

7. The method according to any one of claims 4 to 6, wherein said selection step (e)/(b) is carried out by selecting polyphage comprising the interacting (poly)peptides.

8. The method according to any one of claims 4 to 7, wherein said screenable or selectable property is connected to the infectivity of said RGP.

9. The method according to claim 8, wherein said RGP is encoded by said recombinant vector used in step (a)/(aa) and rendered non-infective and infectivity of said RGP is restored by interaction of said (poly)peptide of step (a)/(aa) with the (poly)peptide of step (b)/(ab) and/or (c)/(ac), said (poly)peptide of step (b)/(ab) and/or (c)/(ac) being fused to a domain that confers infectivity to said RGP.

10. The method according to claim 9, wherein said RGP is rendered non-infective by modification of a genetic sequence which encodes a surface protein necessary for the RGP's binding to and infection of a host cell.

11. The method according to any one of claims 1 to 3, wherein said recombinant vector molecules in step (a)/(aa) give rise to a fusion protein which is expressed on the surface of a cell, preferably a bacterium.

12. The method according to claim 11, wherein said bacterium is Neisseria gonorrhoe or E. coli and said fusion protein consists of at least a part of a flagellum, lam B, peptidoglycan-associated lipoprotein or the Omp A protein and said (poly)peptide.

13. The method according to any one of claims 3 to 7, 11 or 12, wherein said (poly)peptides encoded by said recombinant vector molecules of step (b)/(ab) or (c)/(ac) are linked to at least one screenable or selectable tag.

14. The method according to claim 13, wherein said screenable or selectable tag is encoded by said recombinant vector of step (b)/(ab) or (c)/(ac).

15. The method according to claim 13 or 14, wherein said screenable or selectable tag is selected from the list His(n), myc, FLAG, malE, thioredoxin, GST, streptavidin, beta-galactosidase, alkaline phosphatase, T7 gene 10, Strep-tag and calmodulin.

16. The method according to claim 13, wherein said screenable or selectable tag is encoded by the genome of the host cell.

17. The method according to any one of claims 1 to 16, wherein said (poly)peptides encoded by the nucleic acid sequences of said additional libraries of step (c)/(ac) cause the interaction of said (poly)peptides of steps (a)/(aa) and (b)/(ab) via phosphorylation, glycosylation, methylation, lipidation or farnesylation of at least one of said (poly)peptides of steps (a)/(aa) and (b)/(ab).

18. The method according to any of claims 1 to 10 and 13 to 17, wherein said host cells in step (d)/(a) are spatially addressable, and the nucleic acid sequences mentioned in step (g)/(d) are retrieved from the corresponding spatially addressable host cell.

19. The method according to claim 1 or 2, wherein said screenable or selectable property is expressed intracellularly.

20. The method according to claim 19, wherein said screenable or selectable property is the transactivation of transcription of a reporter gene such as beta-galactosidase, alkaline phosphatase or nutritional markers such as his3 and leu, or resistance genes giving resistance to an antibiotic such as ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

21. The method according to any one of claims 1 to 20, wherein said recombinant vectors of step (a)/(aa), (b)/(ab) and (c)/(ac) comprise recombination promoting sites and in said step (e)/(b) recombination events are selected for, wherein said nucleic acid sequences encoding said (poly)peptides of step (a)/(aa), said nucleic acid sequences encoding said (poly)peptides of step (b)/(ab) and optionally said nucleic acid sequences encoding said (poly)peptides of step (c)/(ac) are contained in the same vector.

22. The method according to claim 21, wherein said recombination events are mediated by the site-specific recombination mechanisms Cre-lox, attP-attB, Mu gin or yeast flp.

23. The method according to claim 21 wherein said recombination promotion sites are restriction enzyme recognition sites and said recombination event is achieved by cutting the recombinant vector molecules mentioned in step (a)/(aa), (b)/(ab) and optionally (c)/(ac) with at least two different restriction enzymes and effecting recombination of the nucleic acid sequences contained in said vectors by ligation.

24. The method according to any one of claims 1 to 23 wherein said identification of said nucleic acid sequences is effected after the selection of step (e)/(b) via PCR and preferably sequencing of said nucleic acid sequences after said PCR.

25. The method according to any one of claims 1 to 24, wherein said recombinant vectors of step (a)/(aa), (b)/(ab) and/or (c)/(ac) comprise at least one gene encoding a selection marker.

26. The method according to claim 25, wherein said selection marker is a resistance to an antibiotic, preferably to ampicillin, chloramphenicol, kanamycin, zeocin, neomycin, tetracycline or streptomycin.

27. The method according to any one of claims 1 to 26, wherein said host cells are F' and preferably E.coli XL-1 Blue, K91 or its derivatives, TG1, XL1kan or TOP10F.

28. The method according to any one of claims 3 to 18 and 21 to 27, wherein said RGPs are produced with the use of helper phage taken from the list R408, M13k07 and VCSM13, M13de13, fCA55 and fKN16 or derivatives thereof.

29. The method according to any of claims 1 to 28, wherein at least one of said genetically diverse nucleic acid sequences encode members of the immunoglobulin superfamily.

30. The method according to claim 29, wherein said genetically diverse nucleic acid sequences encode a repertoire of immunoglobulin heavy or light chains.

31. The method according to any of claims 1 to 30, in which said genetically diverse nucleic acid sequences are generated by a mutagenesis method.

32. The method according to any of claims 1 to 31, in which said genetically diverse nucleic acid sequences are generated from a cDNA library.

33. The method according to any one of claims 1 to 32 wherein said nucleic acid sequences are genes or parts thereof.

34. Kit comprising at least

(a) a recombinant vector molecule as described in step (a)/(aa) or a corresponding vector molecule;

(b) a recombinant vector molecule as described in step (b)/(ab) or a corresponding vector molecule; and, optionally,

(c) at least one further recombinant vector molecule as described in step (c)/(ac) or a corresponding vector molecule.

Figure 1: General description of the polyphage principle

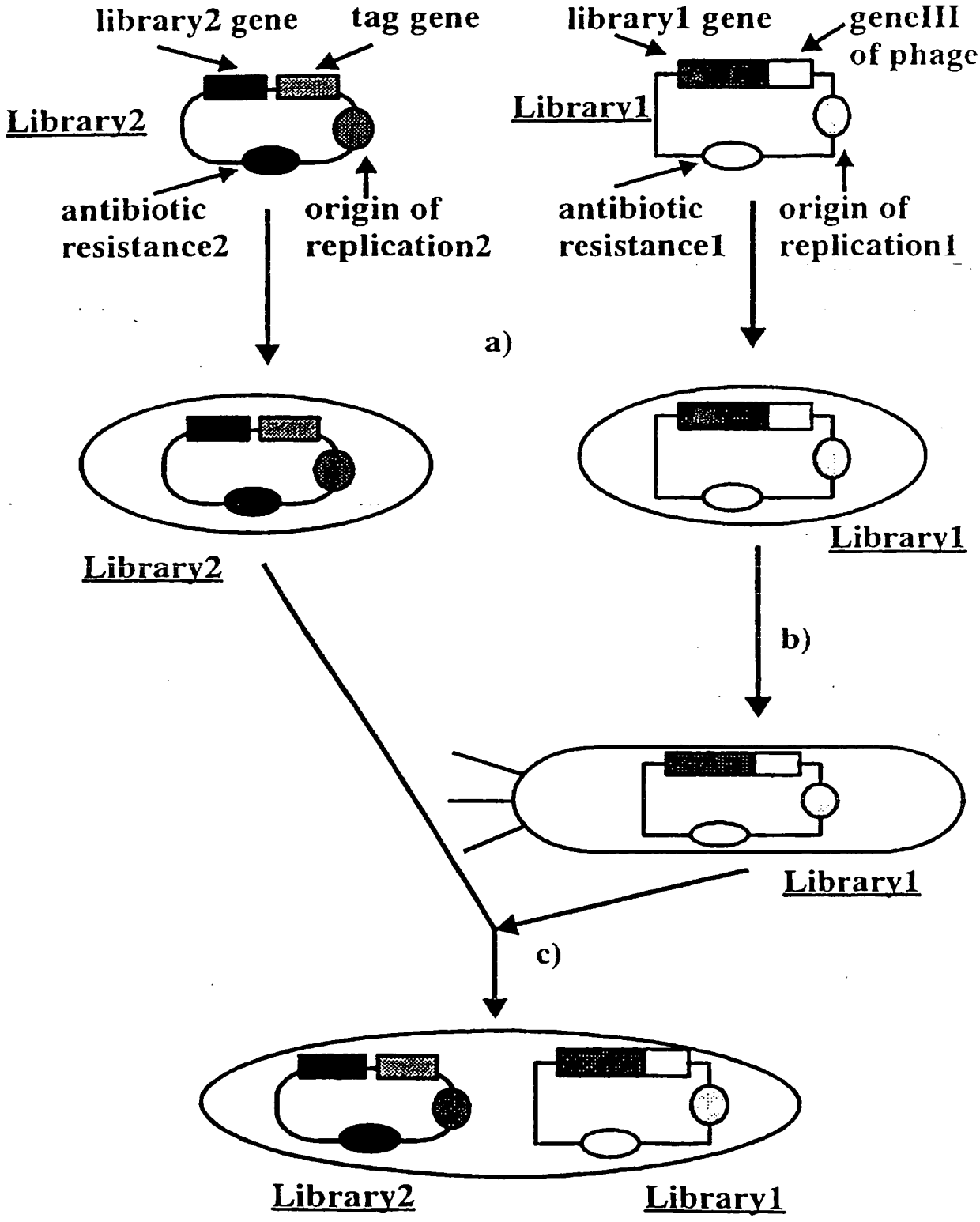


Figure 1: General description of the polyphage principle (cont.)

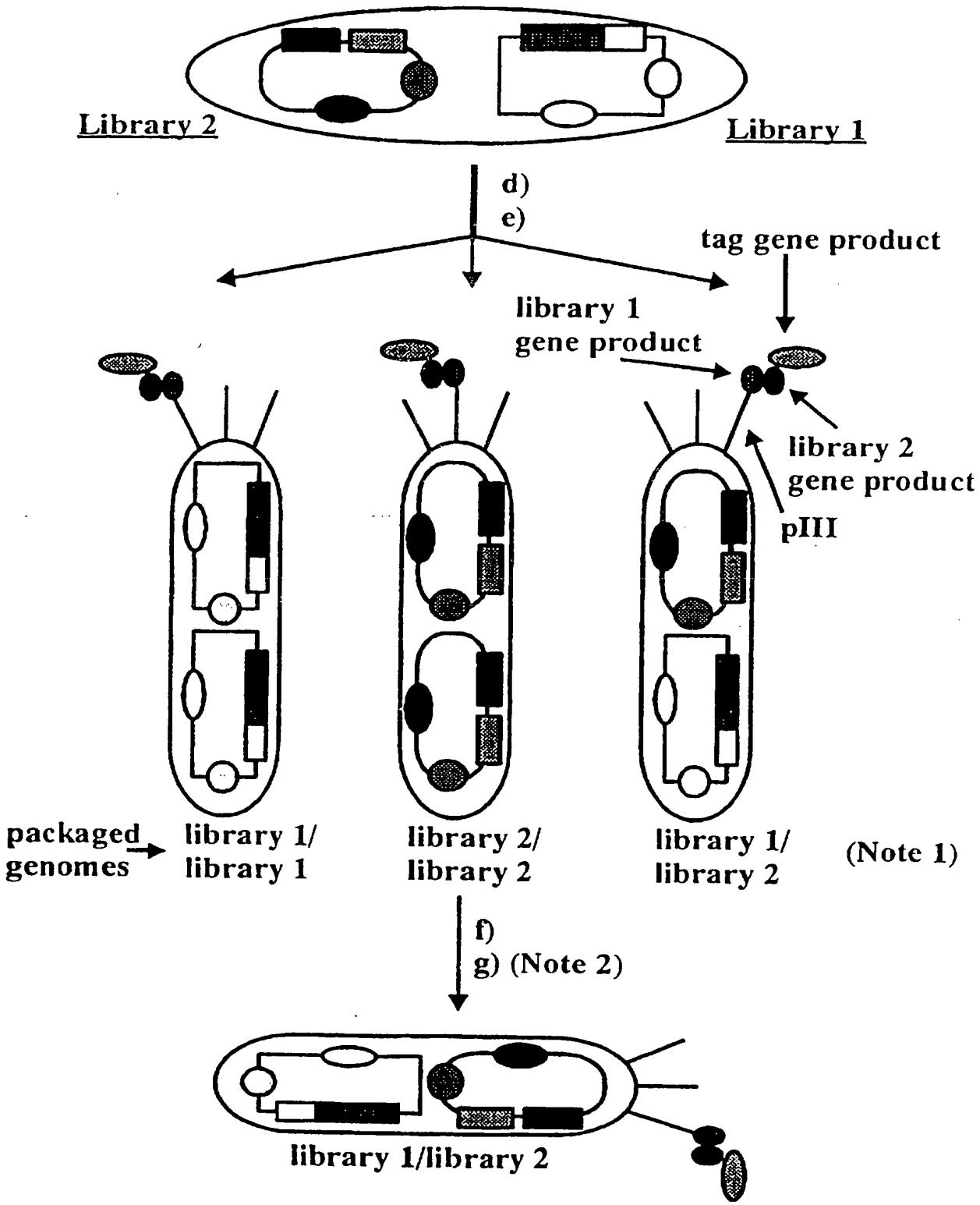


Figure 2: Co-transformation of two phagemids, polyphage formation and selection *via* His-tag: general description

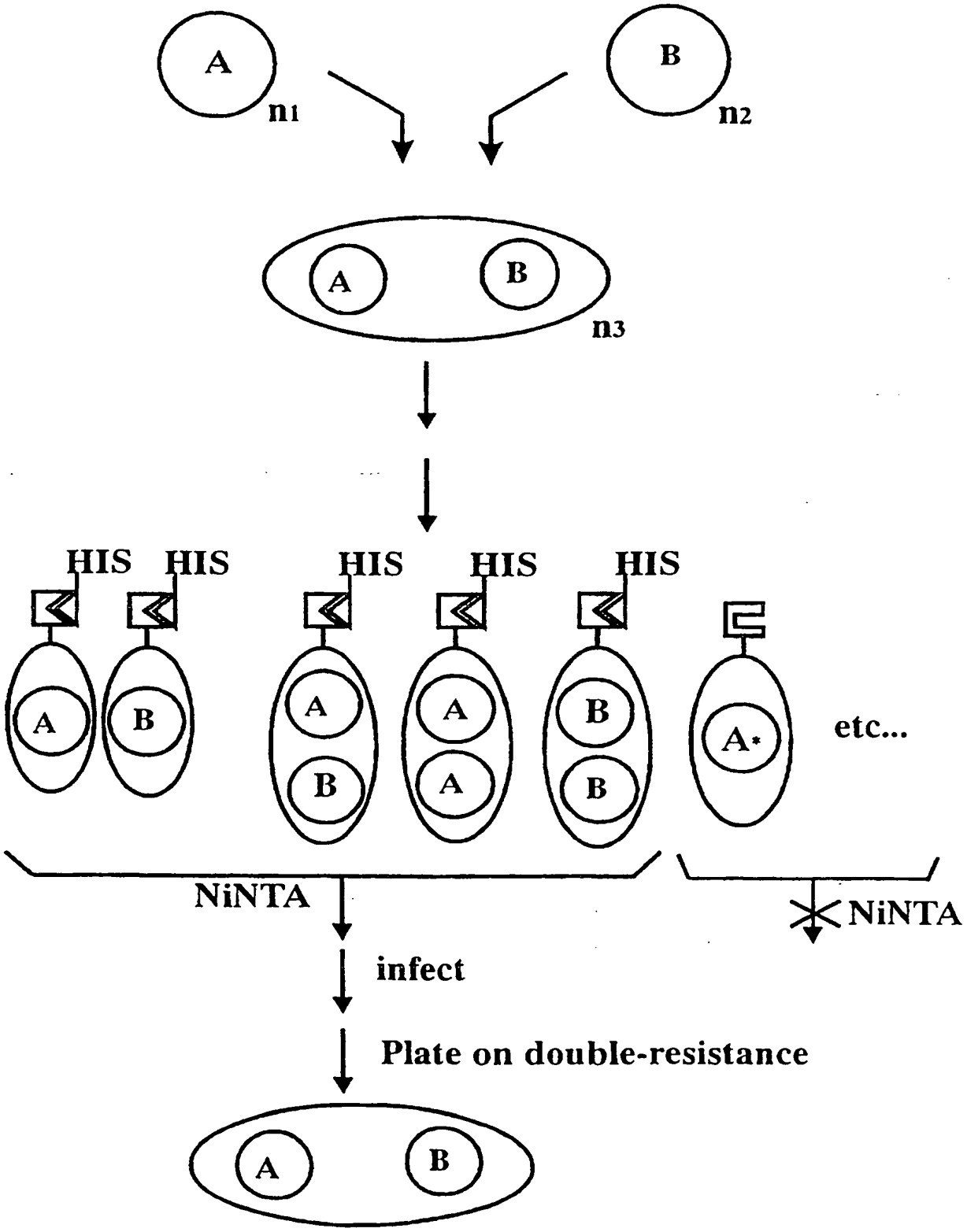


Figure 3: pBS vector series: functional map and sequence of pBS13

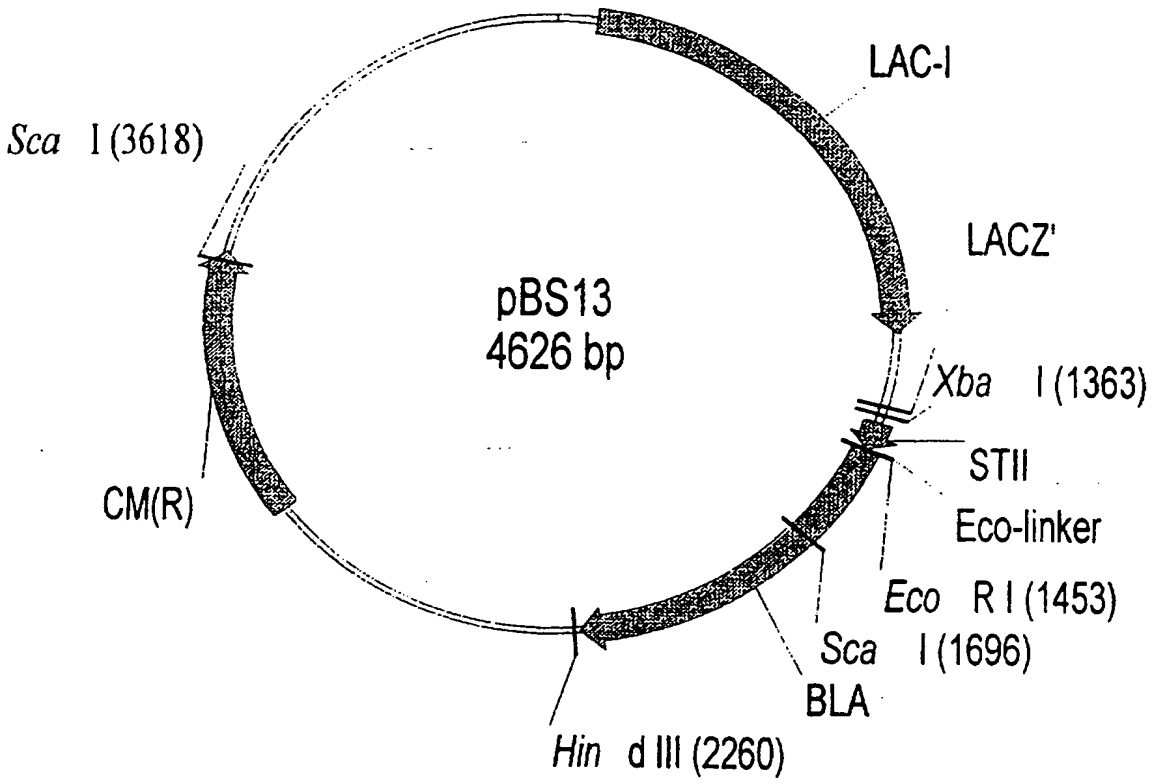


Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
   1  ACCCGACACC ATCGAATGGC GCAAAACCTT TCGCGGTATG GCATGATAGC
      TGGGCTGTGG TAGCTTACCG CGTTTTGGAA AGCGCCATAC CGTACTATCG

  51  GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATGTGAAACC AGTAACGTTA
      CGGGCCTTCT CTCAGTTAAG TCCCACCACT TACACTTTGG TCATTGCAAT

 101  TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
      ATGCTACAGC GTCTCATACG GCCACAGAGA ATAGTCTGGC AAAGGGCGCA

 151  GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAG
      CCACTTGGTC CGGTCGGTGC AAAGACGCTT TTGCGCCCTT TTTCACCTTC

 201  CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG
      GCCGCTACCG CCTCGACTTA ATGTAAGGGT TGGCGCACCG TGTTGTTGAC

 251  GCGGGCAAAC AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT
      CGCCCGTTTG TCAGCAACGA CTAACCGCAA CGGTGGAGGT CAGACCGGGA

 301  GCACGCGCCG TCGCAAATTG TCGCGGCGAT TAAATCTCGC GCCGATCAAC
      CGTGCGCGGC AGCGTTTAAC AGCGCCGCTA ATTTAGAGCG CGGCTAGTTG

 351  TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG CGTCGAAGCC
      ACCCACGGTC GCACCACCAC AGCTACCATC TTGCTTCGCC GCAGCTTCGG

 401  TGTAAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
      ACATTTCGCC GCCACGTGTT AGAAGAGCGC GTTGCGCAGT CACCCGACTA

 451  CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
      GTAATTGATA GGCGACCTAC TGGTCCTACG GTAACGACAC CTTCGACGGA

 501  GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
      CGTGATTACA AGGCCGCAAT AAAGAACTAC AGAGACTGGT CTGTGGGTAG

 551  AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
      TTGTCATAAT AAAAGAGGGT ACTTCTGCCA TGCGCTGACC CGCACCTCGT

 601  TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCCATTAA
      AGACCAGCGT AACCCAGTGG TCGTTTAGCG CGACAATCGC CCGGGTAATT

 651  GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
      CAAGACAGAG CCGCGCAGAC GCAGACCGAC CGACCGTATT TATAGAGTGA

 701  CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
      GCGTTAGTTT AAGTCGGCTA TCGCCTTGCC CTTCCGCTGA CCTCACGGTA

 751  GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
      CAGGCCAAAA GTTGTTTGGT ACGTTTACGA CTTACTCCCG TAGCAAGGGT

 801  CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
      GACGCTACGA CCAACGGTTG CTAGTCTACC GCGACCCGCG TTACGCGCGG
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
 851  ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GACATCTCGG TAGTGGGATA
      TAATGGCTCA GGCCCGACGC GCAACCACGC CTGTAGAGCC ATCACCCTAT

 901  CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTTA ACCACCATCA
      GCTGCTATGG CTTCTGTCGA GTACAATATA GGGCGGCAAT TGGTGGTAGT

 951  AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
      TTGTCCTAAA AGCGGACGAC CCCGTTTGGT CGCACCTGGC GAACGACGTT

1001  CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
      GAGAGAGTCC CGGTCCGCCA CTTCCCGTTA GTCGACAACG GGCAGAGTGA

1051  GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
      CCACTTTTCT TTTTGGTGGG ACCGCGGGTT ATGCGTTTGG CGGAGAGGGG

1101  GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
      CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGGCTGAC

1151  GAAAGCGGGC AGTGAGCGGT ACCCGATAAA AGCGGCTTCC TGACAGGAGG
      CTTTCGCCCG TCACTCGCCA TGGGCTATTT TCGCCGAAGG ACTGTCCTCC

1201  CCGTTTTGTT TTGCAGCCCA CCTCAACGCA ATTAATGTGA GTTAGCTCAC
      GGCAAAACAA AACGTCGGGT GGAGTTGCGT TAATTACACT CAATCGAGTG

1251  TCATTAGGCA CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT
      AGTAATCCGT GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA

1301  GTGGAATTGT GAGCGGATAA CAATTTCACA CAGGAAACAG CTATGACCAT
      CACCTTAACA CTCGCCTATT GTTAAAGTGT GTCCTTTGTC GATACTGGTA

                  XbaI
                  ~~~~~~
1351  GATTACGAAT TTCTAGAGGT TGAGGTGATT TTATGAAAAA GAATATCGCA
      CTAATGCTTA AAGATCTCCA ACTCCACTAA AATACTTTTT CTTATAGCGT

1401  TTTCTTCTTG CATCTATGTT CGTTTTTTCT ATTGCTACAA ATGCATACGC
      AAAGAAGAAC GTAGATACAA GCAAAAAAGA TAACGATGTT TACGTATGCG

        EcoRI
        ~~~~~~
1451  TGAATTCCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
      ACTTAAGGTG GGTCTTTGCG ACCACTTTCA TTTTCTACGA CTTCTAGTCA

1501  TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC
      ACCCACGTGC TCACCCAATG TAGCTTGACC TAGAGTTGTC GCCATTCTAG

1551  CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA
      GAACTCTCAA AAGCGGGGCT TCTTGCAAAA GGTTACTACT CGTGAAAATT

1601  AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC GGGCAAGAGC
      TCAAGACGAT ACACCGCGCC ATAATAGGGC ATAACTGCGG CCCGTTCTCG
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
                                                                  ScaI
                                                                  ~~~~~~
1651  AACTCGGTCG CCGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA
      TTGAGCCAGC GGCGTATGTG ATAAGAGTCT TACTGAACCA ACTCATGAGT

1701  CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG
      GGTCAGTGTC TTTTCGTAGA ATGCCTACCG TACTGTCATT CTCTTAATAC

1751  CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
      GTCACGACGG TATTGGTACT CACTATTGTG ACGCCGGTTG AATGAAGACT

1801  CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG
      GTTGCTAGCC TCCTGGCTTC CTCGATTGGC GAAAAAACGT GTTGTACCCC

1851  GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT
      CTAGTACATT GAGCGGAACT AGCAACCCTT GGCCTCGACT TACTTCGGTA

1901  ACCAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG GCAACAACGT
      TGGTTTGCTG CTCGCACTGT GGTGCTACGG ACATCGTTAC CGTTGTTGCA

1951  TGCGCAAACT ATTAACTGGC GAACTACTTA CTCTAGCTTC CCGGCAACAA
      ACGCGTTTGA TAATTGACCG CTTGATGAAT GAGATCGAAG GGCCGTTGTT

2001  TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC
      AATTATCTGA CCTACCTCCG CCTATTTCAA CGTCCTGGTG AAGACGCGAG

2051  GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
      CCGGGAAGGC CGACCGACCA AATAACGACT ATTTAGACCT CGGCCACTCG

2101  GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC
      CACCCAGAGC GCCATAGTAA CGTCGTGACC CCGGTCTACC ATTCGGGAGG

2151  CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG
      GCATAGCATC AATAGATGTG CTGCCCCTCA GTCCGTTGAT ACCTACTTGC

2201  AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG CATTGGTAAT
      TTTATCTGTC TAGCGACTCT ATCCACGGAG TGACTAATTC GTAACCATTA

            HindIII
            ~~~~~~~
2251  GAGCATGCAA GCTTGACCTG TGAAGTGAAA AATGGCGCAC ATTGTGCGAC
      CTCGTACGTT CGAACTGGAC ACTTCACTTT TTACCGCGTG TAACACGCTG

2301  ATTTTTTTTG TCTGCCGTTT ACCGCTACTG CGTCACGGAT CCCCACGCGC
      TAAAAAAAAC AGACGGCAAA TGGCGATGAC GCAGTGCCTA GGGGTGCGCG

2351  CCTGTAGCGG CGCATTAAGC GCGGCGGGTG TGGTGGTTAC GCGCAGCGTG
      GGACATCGCC GCGTAATTCG CGCCGCCCAC ACCACCAATG CGCGTCGCAC

2401  ACCGCTACAC TTGCCAGCGC CCTAGCGCCC GCTCCTTTCG CTTTCTTCCC
      TGGCGATGTG AACGGTCGCG GGATCGCGGG CGAGGAAAGC GAAAGAAGGG
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
2451  TTCCTTTCTC GCCACGTTCG CCGGCTTTCC CCGTCAAGCT CTAAATCGGG
      AAGGAAAGAG CGGTGCAAGC GGCCGAAAGG GGCAGTTCGA GATTTAGCCC

2501  GCATCCCTTT AGGGTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAAA
      CGTAGGGAAA TCCCAAGGCT AAATCACGAA ATGCCGTGGA GCTGGGGTTT

2551  AAACTTGATT AGGGTGATGG TTCACGTAGT GGGCCATCGC CCTGATAGAC
      TTTGAACTAA TCCCACTACC AAGTGCATCA CCCGGTAGCG GGACTATCTG

2601  GGTTTTTCGC CCTTTGACGT TGGAGTCCAC GTTCTTTAAT AGTGGACTCT
      CCAAAAAGCG GGAAACTGCA ACCTCAGGTG CAAGAAATTA TCACCTGAGA

2651  TGTTCCAAAC TGGAACAACA CTCAACCCTA TCTCGGTCTA TTCTTTTGAT
      ACAAGGTTTG ACCTTGTTGT GAGTTGGGAT AGAGCCAGAT AAGAAAACTA

2701  TTATAAGGGA TTTTGCCGAT TTCGGCCTAT TGGTTAAAAA ATGAGCTGAT
      AATATTCCCT AAAACGGCTA AAGCCGGATA ACCAATTTTT TACTCGACTA

2751  TTAACAAAAA TTTAACGCGA ATTTTAACAA AATATTAACG TTTACAATTT
      AATTGTTTTT AAATTGCGCT TAAAATTGTT TTATAATTGC AAATGTTAAA

2801  CAGGTGGCAC TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT
      GTCCACCGTG AAAAGCCCCT TTACACGCGC CTTGGGGATA AACAAATAAA

2851  TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGTCGAGAC GTTGGGTGAG
      AAGATTTATG TAAGTTTATA CATAGGCGAG TACAGCTCTG CAACCCACTC

2901  GTTCCAACTT TCACCATAAT GAAATAAGAT CACTACCGGG CGTATTTTTT
      CAAGGTTGAA AGTGGTATTA CTTTATTCTA GTGATGGCCC GCATAAAAAA

2951  GAGTTATCGA GATTTTCAGG AGCTAAGGAA GCTAAAATGG AGAAAAAAAT
      CTCAATAGCT CTAAAAGTCC TCGATTCCTT CGATTTTACC TCTTTTTTTA

3001  CACTGGATAT ACCACCGTTG ATATATCCCA ATGGCATCGT AAAGAACATT
      GTGACCTATA TGGTGGCAAC TATATAGGGT TACCGTAGCA TTTCTTGTAA

3051  TTGAGGCATT TCAGTCAGTT GCTCAATGTA CCTATAACCA GACCGTTCAG
      AACTCCGTAA AGTCAGTCAA CGAGTTACAT GGATATTGGT CTGGCAAGTC

3101  CTGGATATTA CGGCCTTTTT AAAGACCGTA AAGAAAAATA AGCACAAGTT
      GACCTATAAT GCCGGAAAAA TTTCTGGCAT TTCTTTTTAT TCGTGTTCAA

3151  TTATCCGGCC TTTATTCACA TTCTTGCCCG CCTGATGAAT GCTCATCCGG
      AATAGGCCGG AAATAAGTGT AAGAACGGGC GGACTACTTA CGAGTAGGCC

3201  AGTTCCGTAT GGCAATGAAA GACGGTGAGC TGGTGATATG GGATAGTGTT
      TCAAGGCATA CCGTTACTTT CTGCCACTCG ACCACTATAC CCTATCACAA

3251  CACCCTTGTT ACACCGTTTT CCATGAGCAA ACTGAAACGT TTTCATCGCT
      GTGGGAACAA TGTGGCAAAA GGTACTCGTT TGACTTTGCA AAAGTAGCGA
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
3301  CTGGAGTGAA TACCACGACG ATTTCCGGCA GTTTCTACAC ATATATTCGC
      GACCTCACTT ATGGTGCTGC TAAAGGCCGT CAAAGATGTG TATATAAGCG

3351  AAGATGTGGC GTGTTACGGT GAAAACCTGG CCTATTTCCC TAAAGGGTTT
      TTCTACACCG CACAATGCCA CTTTTGGACC GGATAAAGGG ATTTCCCAAA

3401  ATTGAGAATA TGTTTTTCGT CTCAGCCAAT CCCTGGGTGA GTTTCACCAG
      TAACTCTTAT ACAAAAAGCA GAGTCGGTTA GGGACCCACT CAAAGTGGTC

3451  TTTTGATTTA AACGTGGCCA ATATGGACAA CTTCTTCGCC CCCGTTTTCA
      AAAACTAAAT TTGCACCGGT TATACCTGTT GAAGAAGCGG GGGCAAAAGT

3501  CCATGGGCAA ATATTATACG CAAGGCGACA AGGTGCTGAT GCCGCTGGCG
      GGTACCCGTT TATAATATGC GTTCCGCTGT TCCACGACTA CGGCGACCGC

3551  ATTCAGGTTC ATCATGCCGT CTGTGATGGC TTCCATGTCG GCAGAATGCT
      TAAGTCCAAG TAGTACGGCA GACACTACCG AAGGTACAGC CGTCTTACGA

                      ScaI
                      ~~~~~~
3601  TAATGAATTA CAACAGTACT GCGATGAGTG GCAGGGCGGG GCGTAATTTT
      ATTACTTAAT GTTGTCATGA CGCTACTCAC CGTCCCGCCC CGCATTAAAA

3651  TTTAAGGCAG TTATTGGTGC CCTTAAACGC CTGGTGCTAC GCCTGAATAA
      AAATTCCGTC AATAACCACG GGAATTTGCG GACCACGATG CGGACTTATT

3701  GTGATAATAA GCGGATGAAT GGCAGAAATT CGAAAGCAAA TTCGACCCGG
      CACTATTATT CGCCTACTTA CCGTCTTTAA GCTTTCGTTT AAGCTGGGCC

3751  TCGTCGGTTC AGGGCAGGGT CGTTAAATAG CCGCTTATGT CTATTGCTGG
      AGCAGCCAAG TCCCGTCCCA GCAATTTATC GGCGAATACA GATAACGACC

3801  TTTACCGGTT TATTGACTAC CGGAAGCAGT GTGACCGTGT GCTTCTCAAA
      AAATGGCCAA ATAACTGATG GCCTTCGTCA CACTGGCACA CGAAGAGTTT

3851  TGCCTGAGGC CAGTTTGCTC AGGCTCTCCC CGTGGAGGTA ATAATTGCTC
      ACGGACTCCG GTCAAACGAG TCCGAGAGGG GCACCTCCAT TATTAACGAG

3901  GACATGACCA AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA
      CTGTACTGGT TTTAGGGAAT TGCACTCAAA AGCAAGGTGA CTCGCAGTCT

3951  CCCCGTAGAA AAGATCAAAG GATCTTCTTG AGATCCTTTT TTTCTGCGCG
      GGGGCATCTT TTCTAGTTTC CTAGAAGAAC TCTAGGAAAA AAAGACGCGC

4001  TAATCTGCTG CTTGCAAACA AAAAAACCAC CGCTACCAGC GGTGGTTTGT
      ATTAGACGAC GAACGTTTGT TTTTTTGGTG GCGATGGTCG CCACCAAACA

4051  TTGCCGGATC AAGAGCTACC AACTCTTTTT CCGAAGGTAA CTGGCTTCAG
      AACGGCCTAG TTCTCGATGG TTGAGAAAAA GGCTTCCATT GACCGAAGTC
```

Figure 3: pBS vector series: functional map and sequence of pBS13 (continued)

```
4101  CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG TAGTTAGGCC
      GTCTCGCGTC TATGGTTTAT GACAGGAAGA TCACATCGGC ATCAATCCGG

4151  ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC
      TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG AGACGATTAG

4201  CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT
      GACAATGGTC ACCGACGACG GTCACCGCTA TTCAGCACAG AATGGCCCAA

4251  GGACTCAAGA CGATAGTTAC CGGATAAGGC GCAGCGGTCG GGCTGAACGG
      CCTGAGTTCT GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC

4301  GGGGTTCGTG CACACAGCCC AGCTTGGAGC GAACGACCTA CACCGAACTG
      CCCCAAGCAC GTGTGTCGGG TCGAACCTCG CTTGCTGGAT GTGGCTTGAC

4351  AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC CCGAAGGGAG
      TCTATGGATG TCGCACTCGA TACTCTTTCG CGGTGCGAAG GGCTTCCCTC

4401  AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA
      TTTCCGCCTG TCCATAGGCC ATTCGCCGTC CCAGCCTTGT CCTCTCGCGT

4451  CGAGGGAGCT TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG
      GCTCCCTCGA AGGTCCCCCT TTGCGGACCA TAGAAATATC AGGACAGCCC

4501  TTTCGCCACC TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG
      AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA GCAGTCCCCC

4551  GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCCTTTTTA CGGTTCCTGG
      CGCCTCGGAT ACCTTTTTGC GGTCGTTGCG CCGGAAAAAT GCCAAGGACC

4601  CCTTTTGCTG GCCTTTTGCT CACATG
      GGAAAACGAC CGGAAAACGA GTGTAC
```

Figure 4: Co-existence of phagemids: results of restriction digest

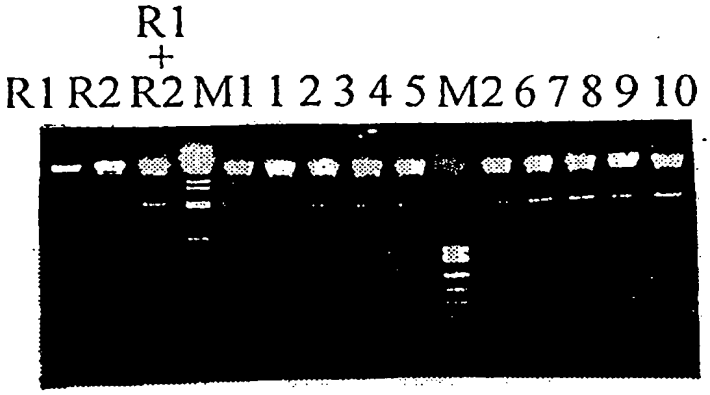


Figure 5: Phagemid vector pYING1-C1: functional map

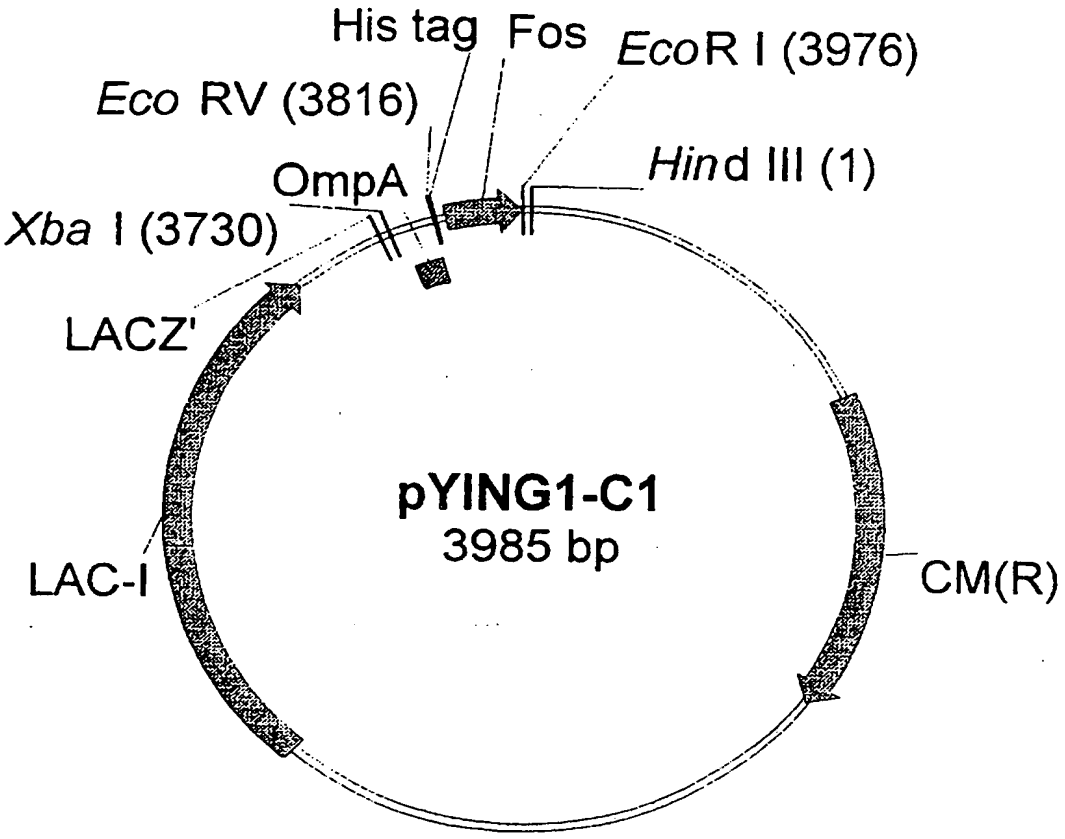


Figure 6: Phagemid vector pYANG3-A: functional map

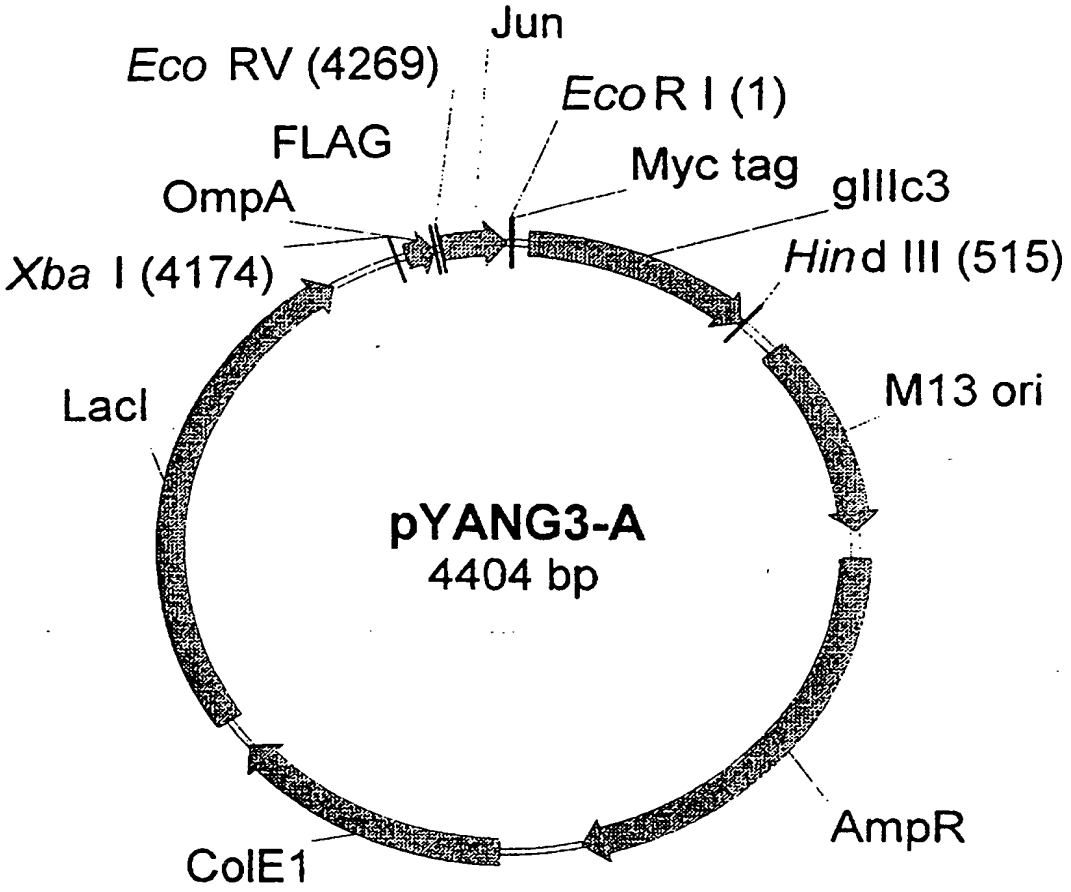
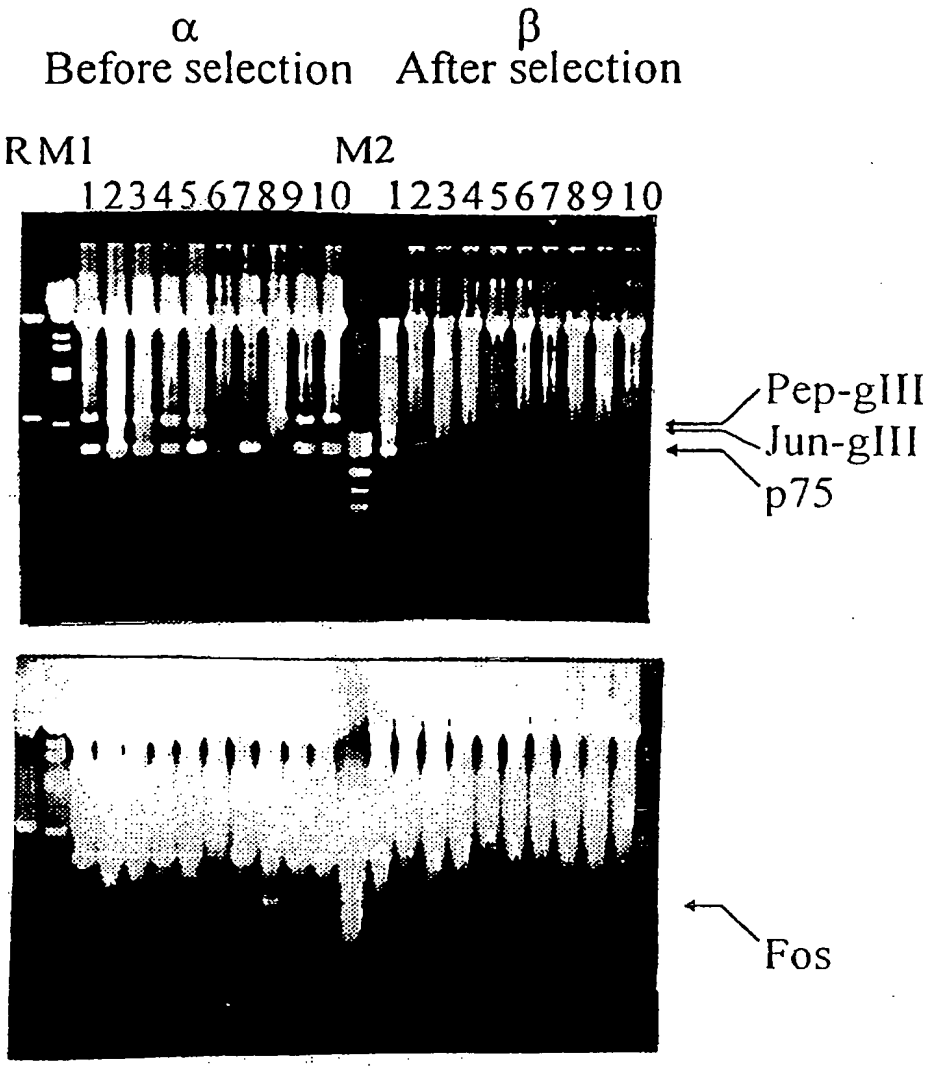


Figure 7: Analysis of selected clones (see Table 2)

7.a: Restriction digest of clones before and after selection



7.b: PCR of clones after selection with primers OPEP5L and OGIII3

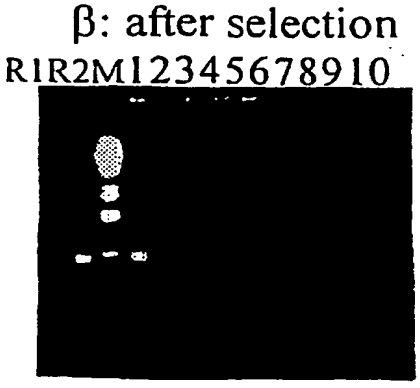


Figure 8: Phagemid vector pING1-C1: functional map

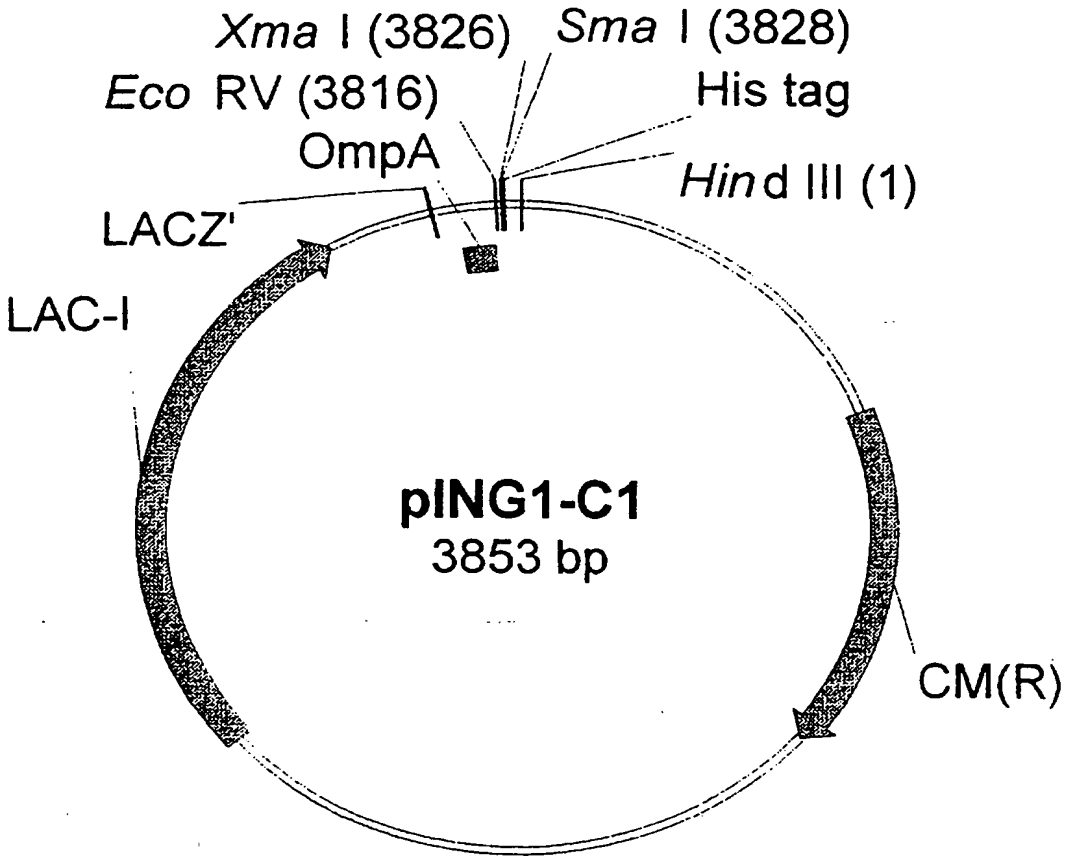


Figure 9: Phagemid vector pONG3-A: functional map

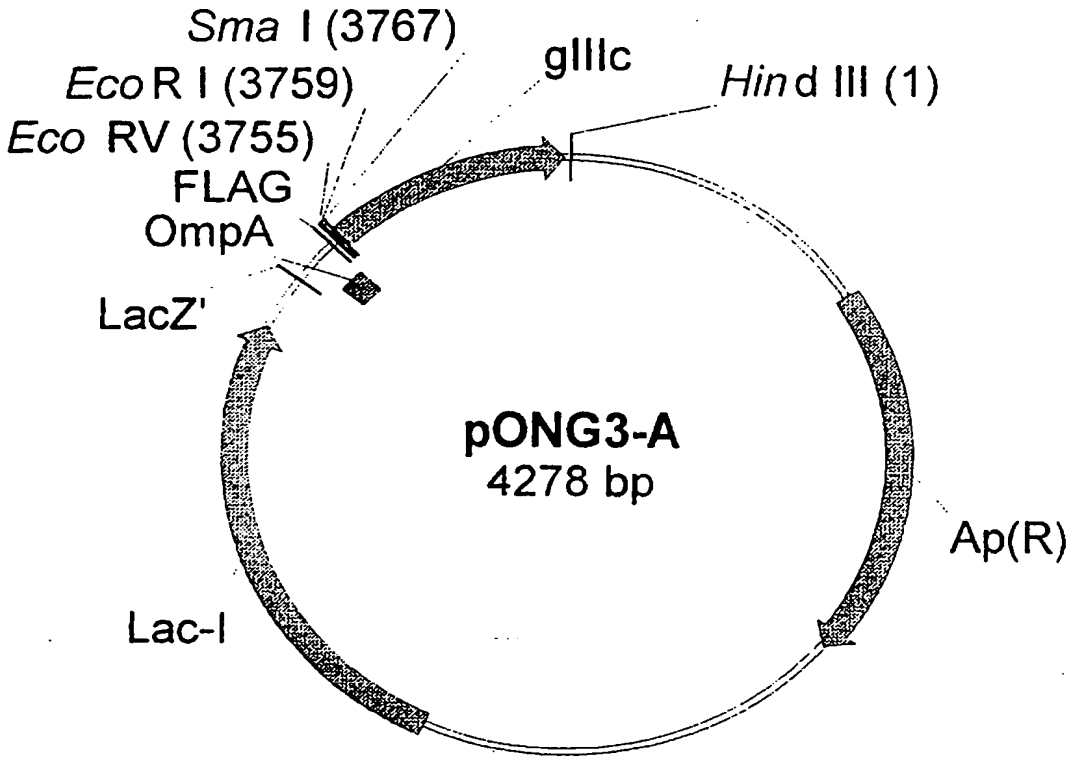


Figure 10: Co-transformation of phage and plasmid, polyphage formation and selection *via* SIP: general description

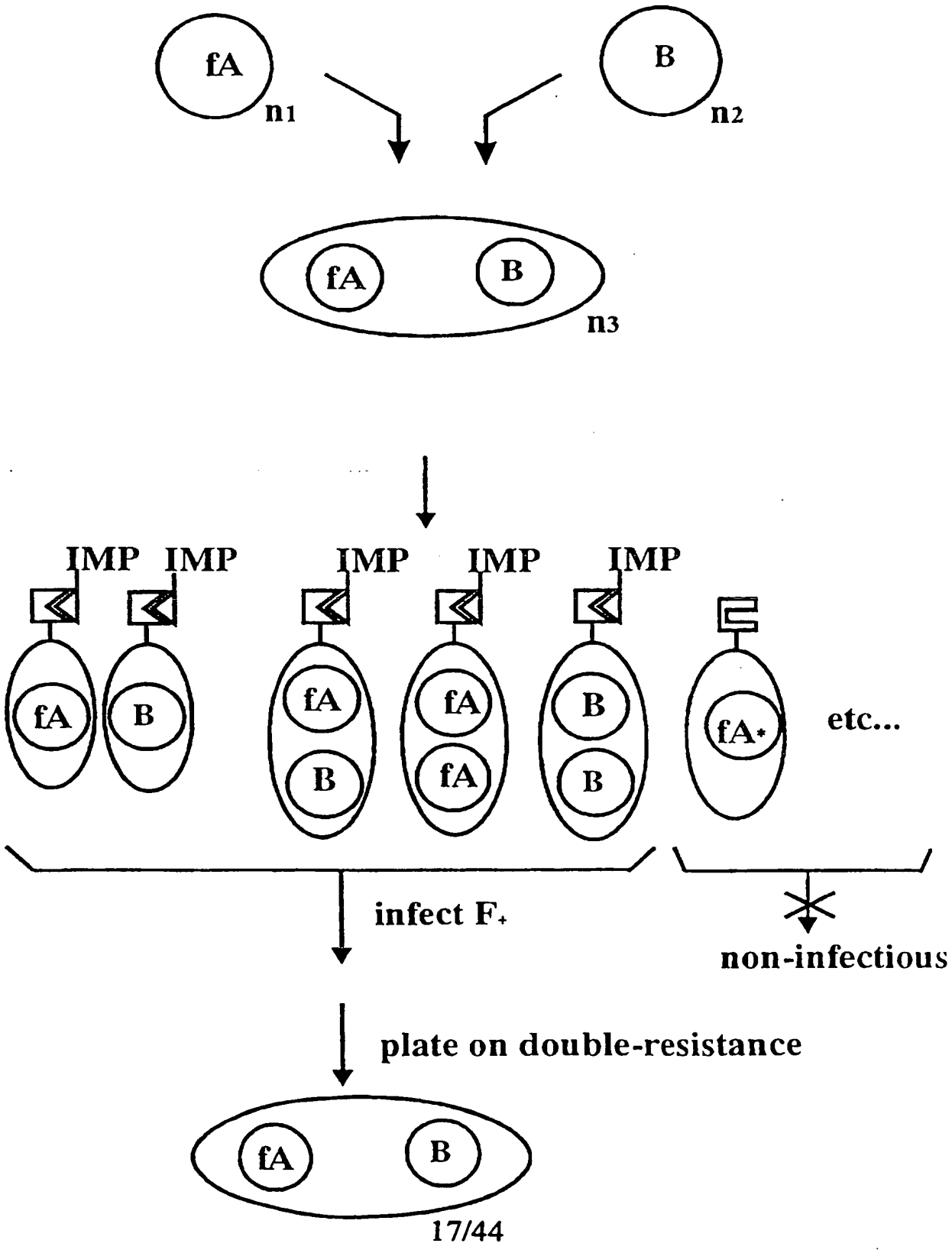


Figure 11: Phage vector fhag1A: functional map

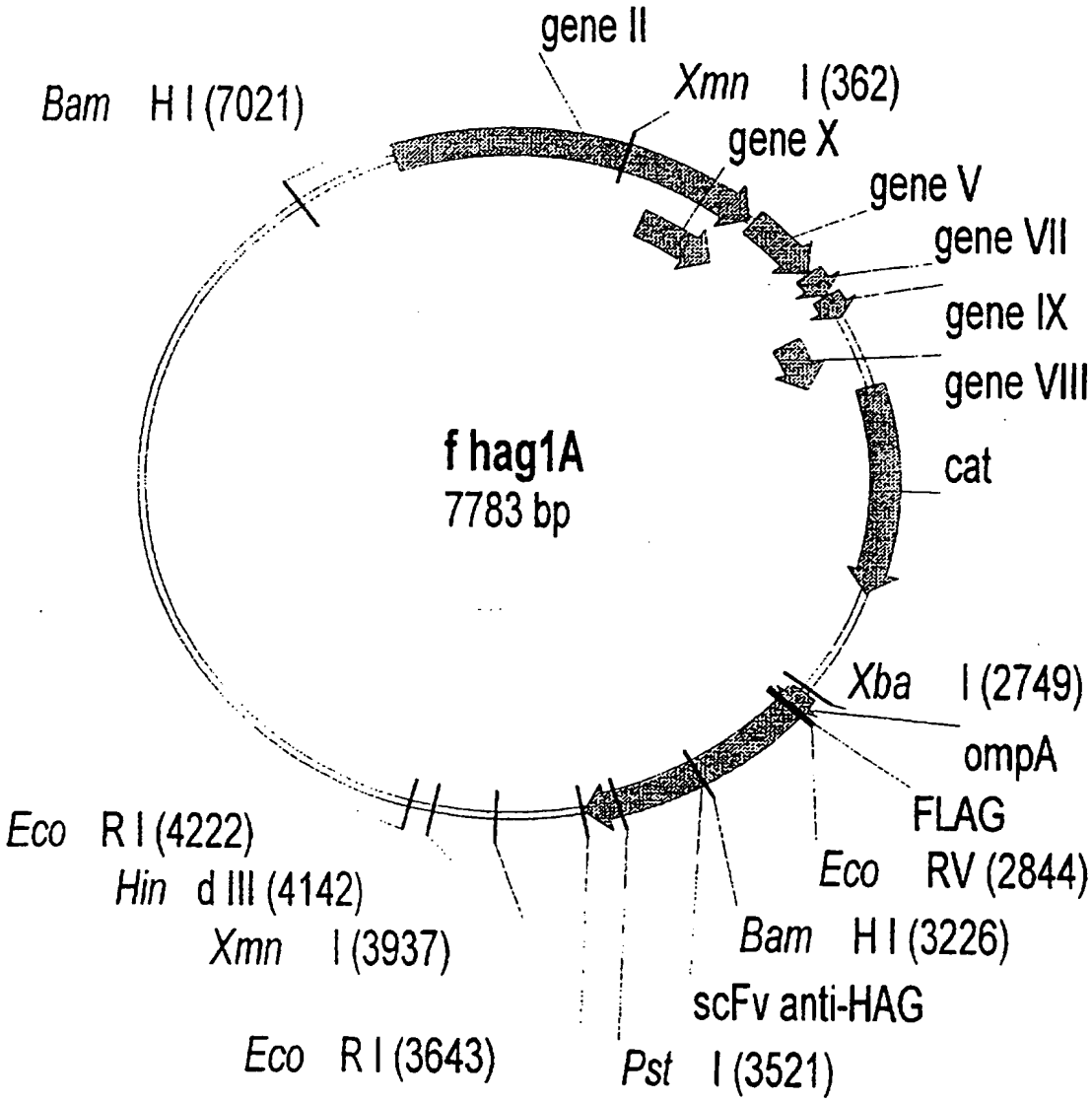


Figure 11a: CAT gene module: functional map and sequence

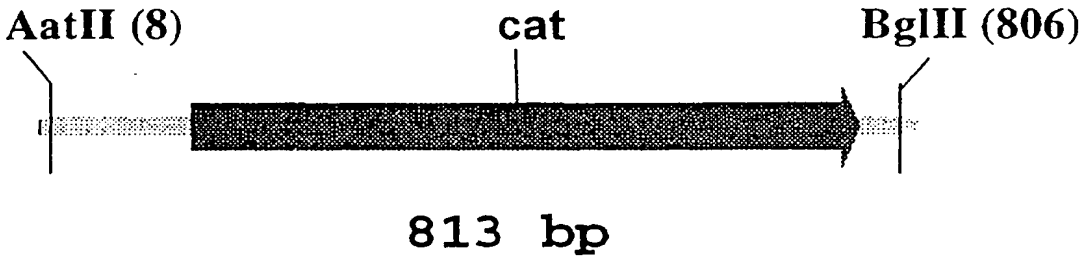


Figure 11a: CAT gene module: functional map and sequence (cont.)

```
          AatII
          ------
     1  GGGACGTCGG GTGAGGTTCC AACTTTCACC ATAATGAAAT AAGATCACTA
        CCCTGCAGCC CACTCCAAGG TTGAAAGTGG TATTACTTTA TTCTAGTGAT

    51  CCGGGCGTAT TTTTTGAGTT ATCGAGATTT TCAGGAGCTA AGGAAGCTAA
        GGCCCGCATA AAAAACTCAA TAGCTCTAAA AGTCCTCGAT TCCTTCGATT

   101  AATGGAGAAA AAAATCACTG GATATACCAC CGTTGATATA TCCCAATGGC
        TTACCTCTTT TTTTAGTGAC CTATATGGTG GCAACTATAT AGGGTTACCG

   151  ATCGTAAAGA ACATTTTGAG GCATTTCAGT CAGTTGCTCA ATGTACCTAT
        TAGCATTTCT TGTAAAACTC CGTAAAGTCA GTCAACGAGT TACATGGATA

   201  AACCAGACCG TTCAGCTGGA TATTACGGCC TTTTTAAAGA CCGTAAAGAA
        TTGGTCTGGC AAGTCGACCT ATAATGCCGG AAAAATTTCT GGCATTTCTT

   251  AAATAAGCAC AAGTTTTATC CGGCCTTTAT TCACATTCTT GCCCGCCTGA
        TTTATTCGTG TTCAAAATAG GCCGGAAATA AGTGTAAGAA CGGGCGGACT

   301  TGAATGCTCA CCCGGAGTTC CGTATGGCAA TGAAAGACGG TGAGCTGGTG
        ACTTACGAGT GGGCCTCAAG GCATACCGTT ACTTTCTGCC ACTCGACCAC

   351  ATATGGGATA GTGTTCACCC TTGTTACACC GTTTTCCATG AGCAAACTGA
        TATACCCTAT CACAAGTGGG AACAATGTGG CAAAAGGTAC TCGTTTGACT

   401  AACGTTTTCA TCGCTCTGGA GTGAATACCA CGACGATTTC CGGCAGTTTC
        TTGCAAAAGT AGCGAGACCT CACTTATGGT GCTGCTAAAG GCCGTCAAAG

   451  TACACATATA TTCGCAAGAT GTGGCGTGTT ACGGTGAAAA CCTGGCCTAT
        ATGTGTATAT AAGCGTTCTA CACCGCACAA TGCCACTTTT GGACCGGATA

   501  TTCCCTAAAG GGTTTATTGA GAATATGTTT TTCGTCTCAG CCAATCCCTG
        AAGGGATTTC CCAAATAACT CTTATACAAA AAGCAGAGTC GGTTAGGGAC

   551  GGTGAGTTTC ACCAGTTTTG ATTTAAACGT AGCCAATATG GACAACTTCT
        CCACTCAAAG TGGTCAAAAC TAAATTTGCA TCGGTTATAC CTGTTGAAGA

   601  TCGCCCCCGT TTTCACTATG GGCAAATATT ATACGCAAGG CGACAAGGTG
        AGCGGGGGCA AAAGTGATAC CCGTTTATAA TATGCGTTCC GCTGTTCCAC

   651  CTGATGCCGC TGGCGATTCA GGTTCATCAT GCCGTTTGTG ATGGCTTCCA
        GACTACGGCG ACCGCTAAGT CCAAGTAGTA CGGCAAACAC TACCGAAGGT

   701  TGTCGGCAGA ATGCTTAATG AATTACAACA GTACTGCGAT GAGTGGCAGG
        ACAGCCGTCT TACGAATTAC TTAATGTTGT CATGACGCTA CTCACCGTCC

   751  GCGGGGCGTA ATTTTTTTAA GGCAGTTATT GGGTGCCCTT AAACGCCTGG
        CGCCCCGCAT TAAAAAAATT CCGTCAATAA CCCACGGGAA TTTGCGGACC

          BglII
          ------
   801  TGCTAGATCT TCC
        ACGATCTAGA AGG
```

Figure 12: Phage vector fjun1A: functional map

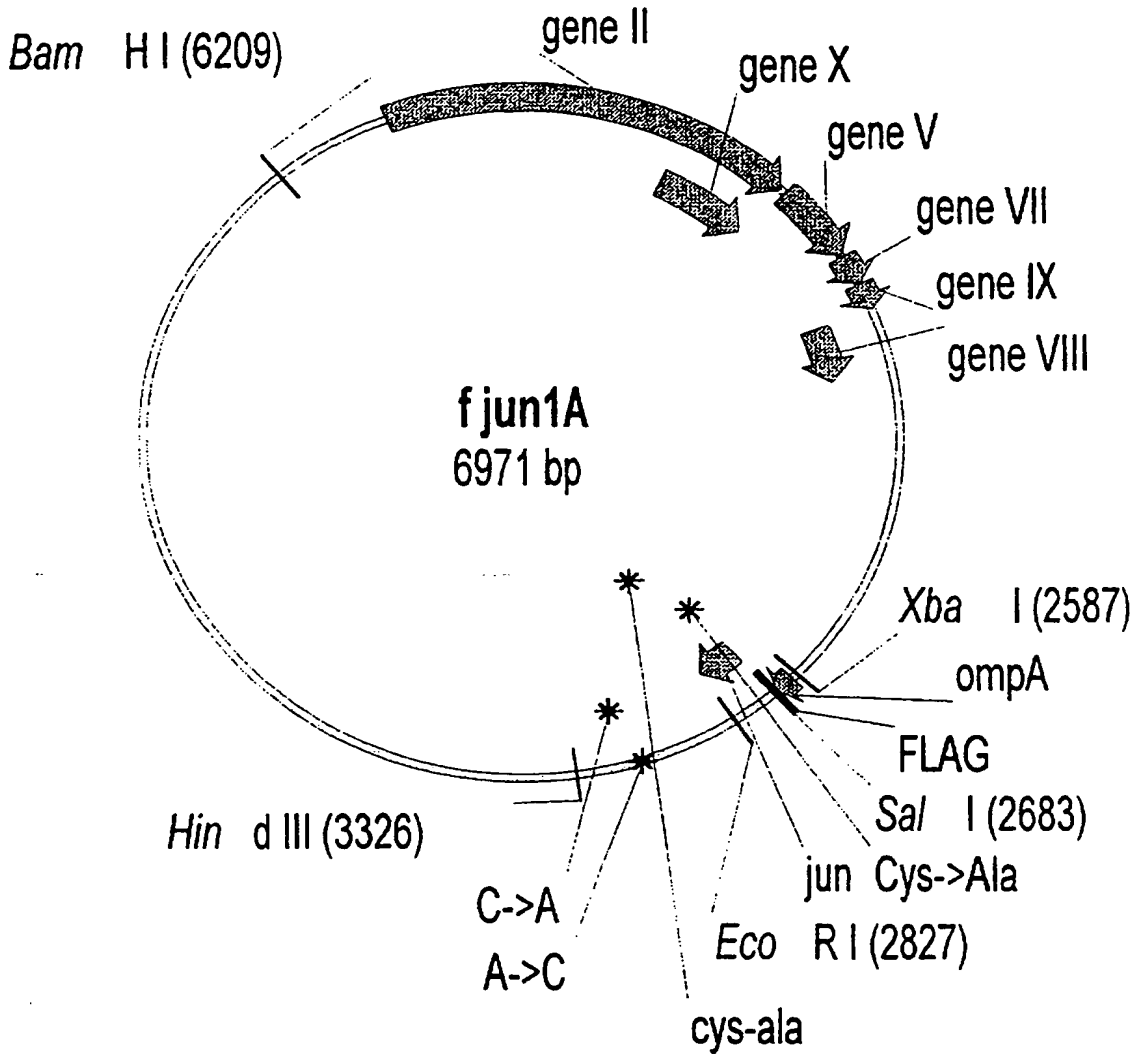


Figure 13: Phage vector fjun1B: functional map

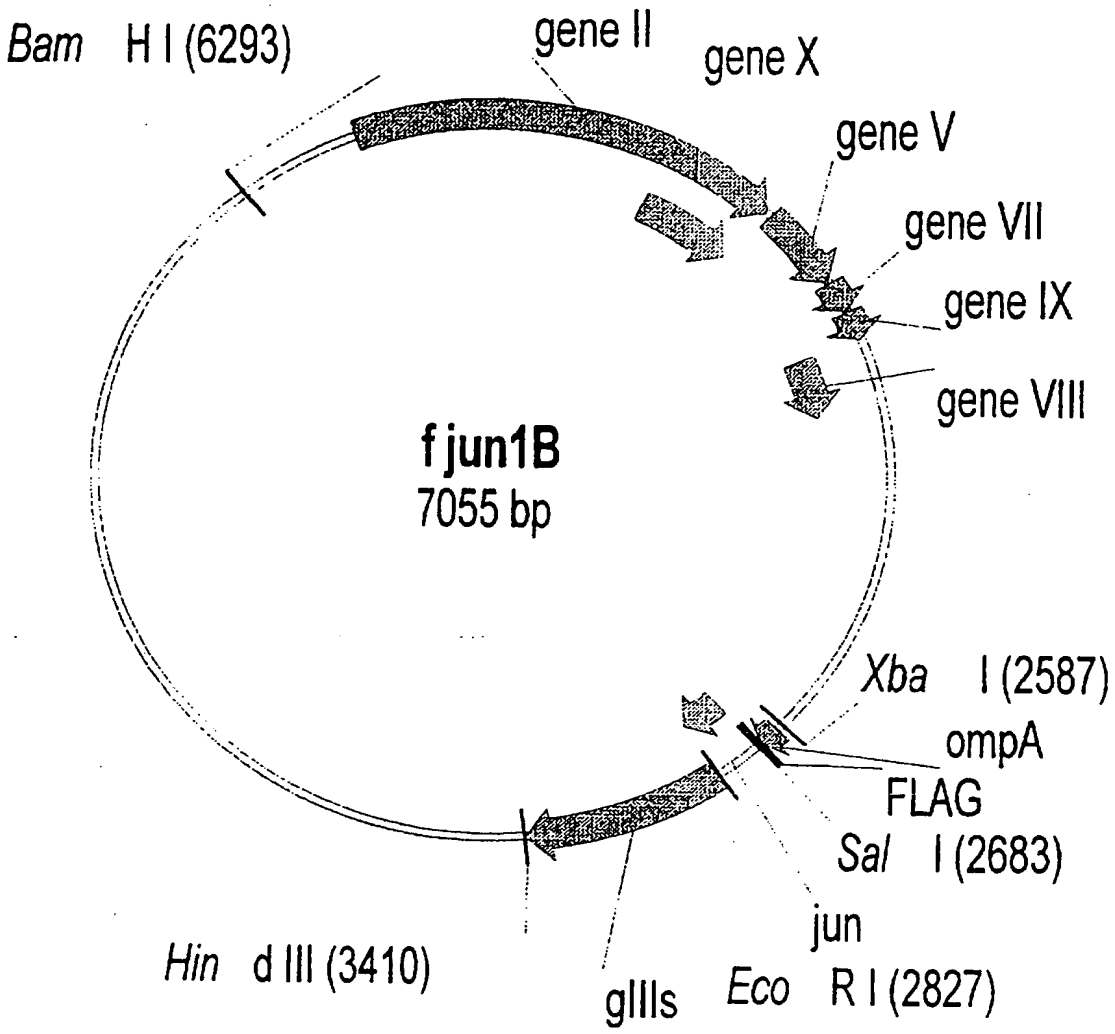


Figure 14: Phage vector fpep3_1B: functional map

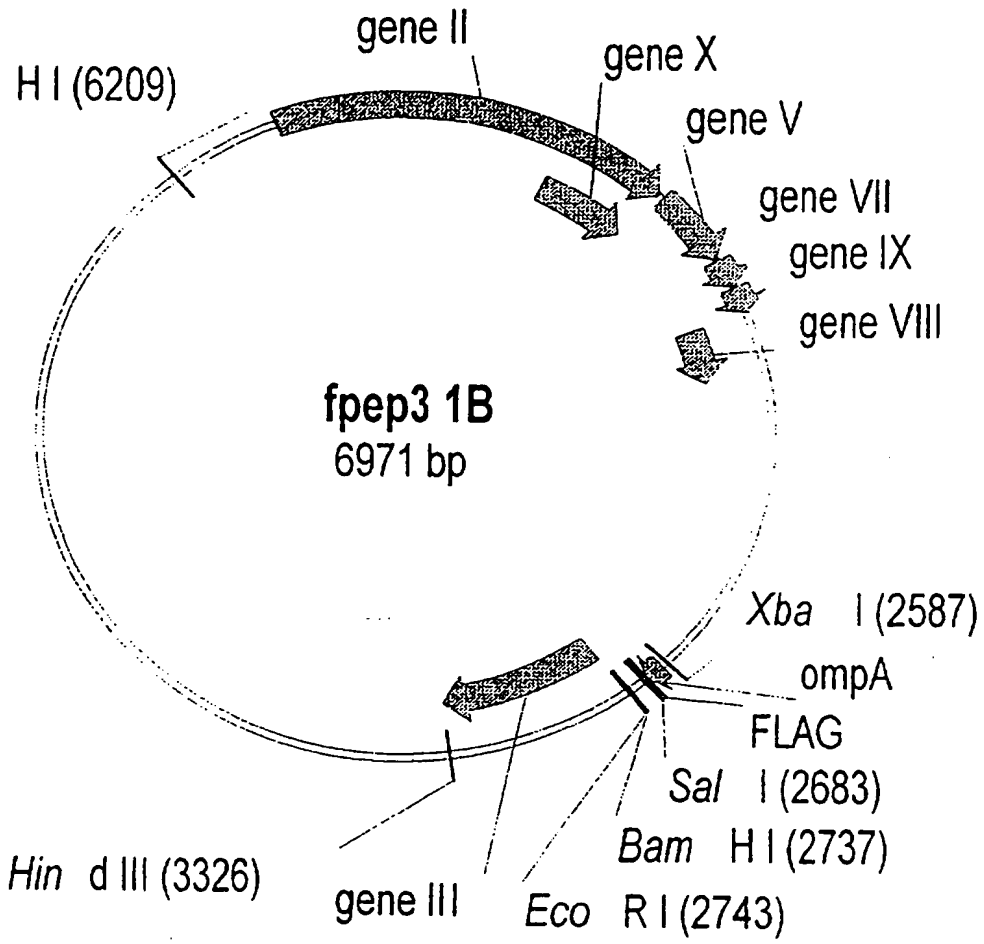


Figure 15: Phage vector fNGF_1B: functional map

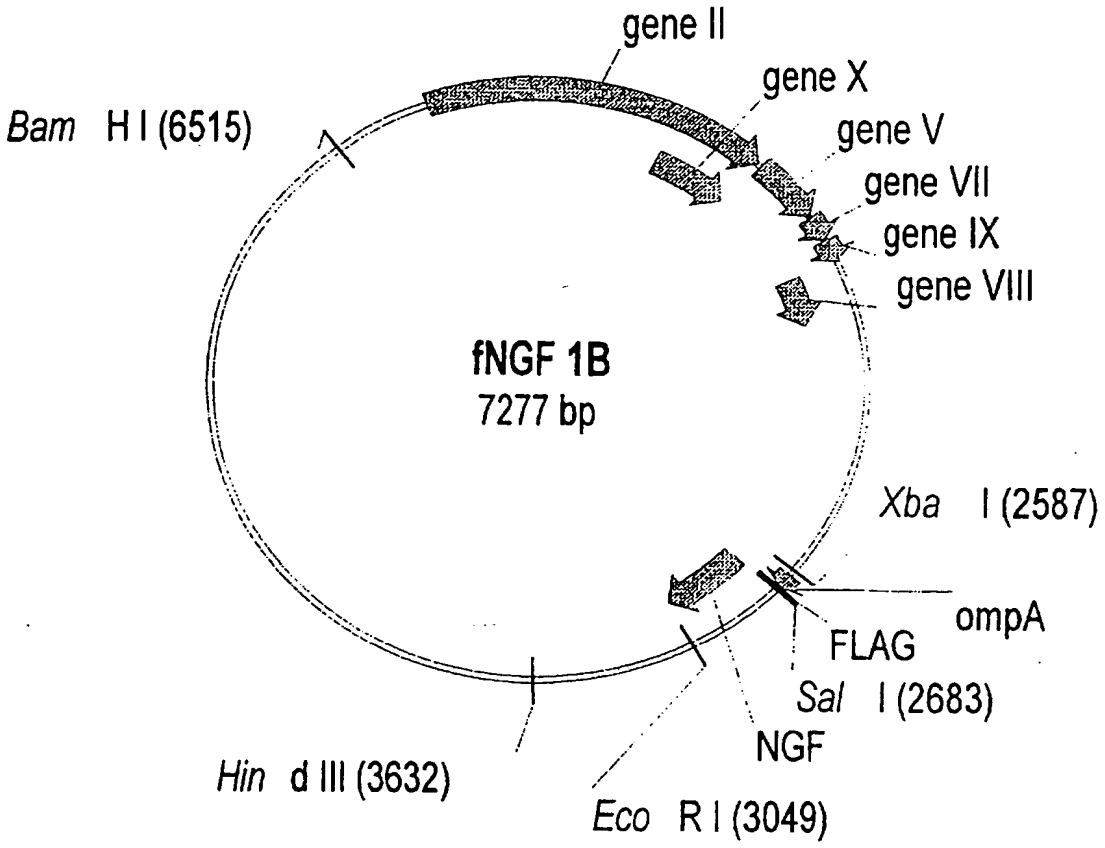


Figure 16: Plasmid pUC19/IMPhag: functional map

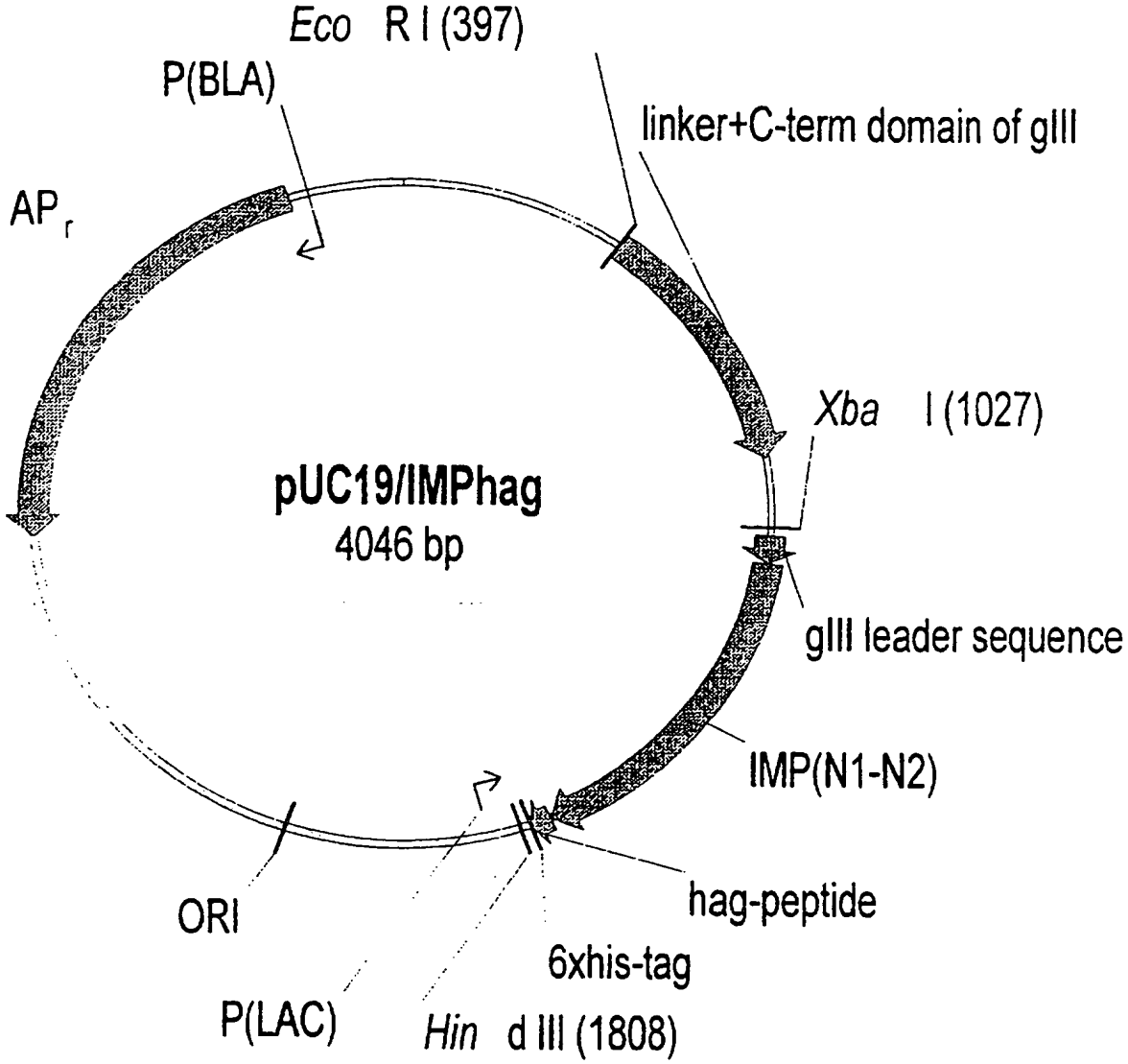


Figure 17: Plasmid pUC18/IMPp75: functional map

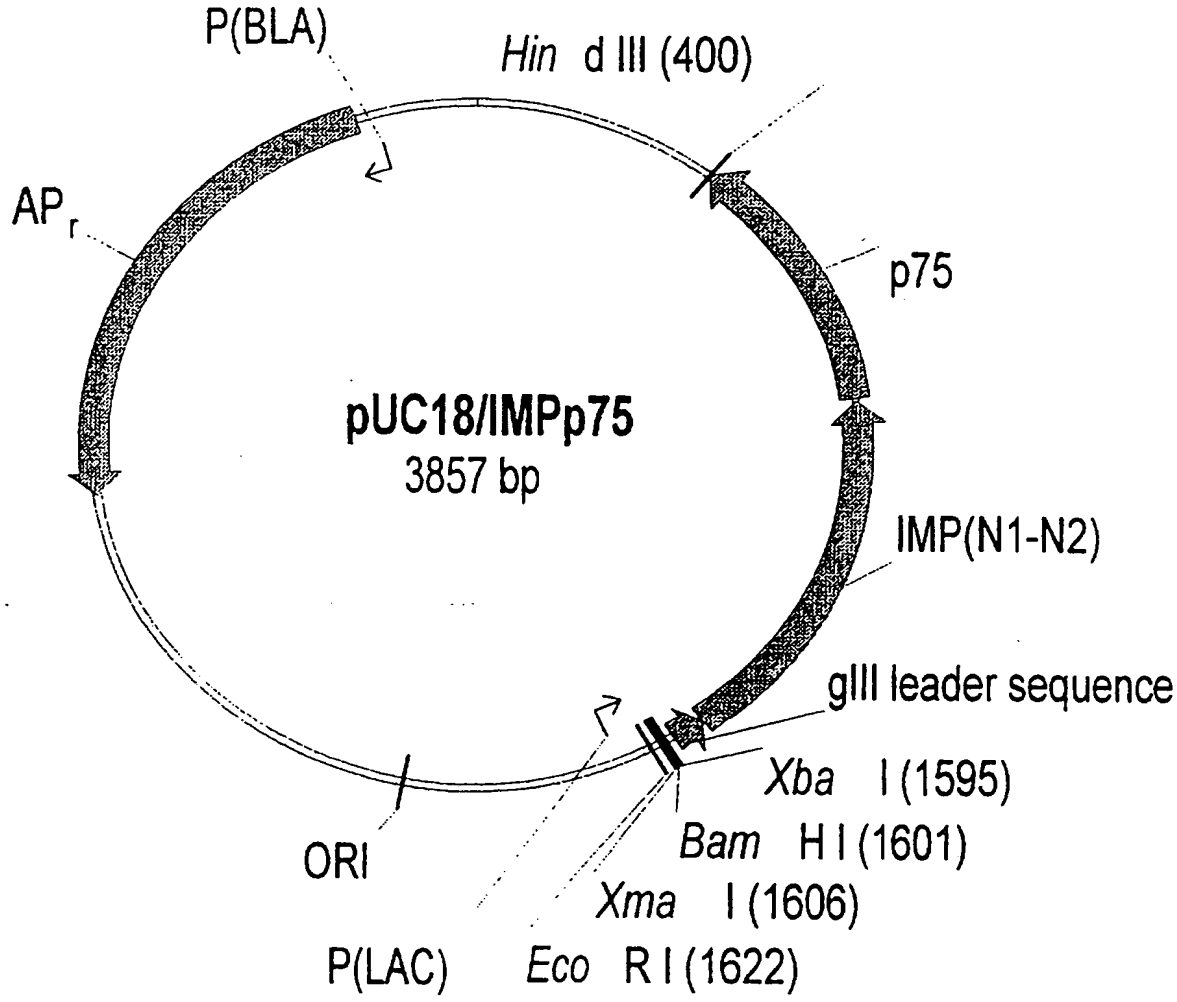


Figure 18: Plasmid pUC18/IMPIL16: functional map

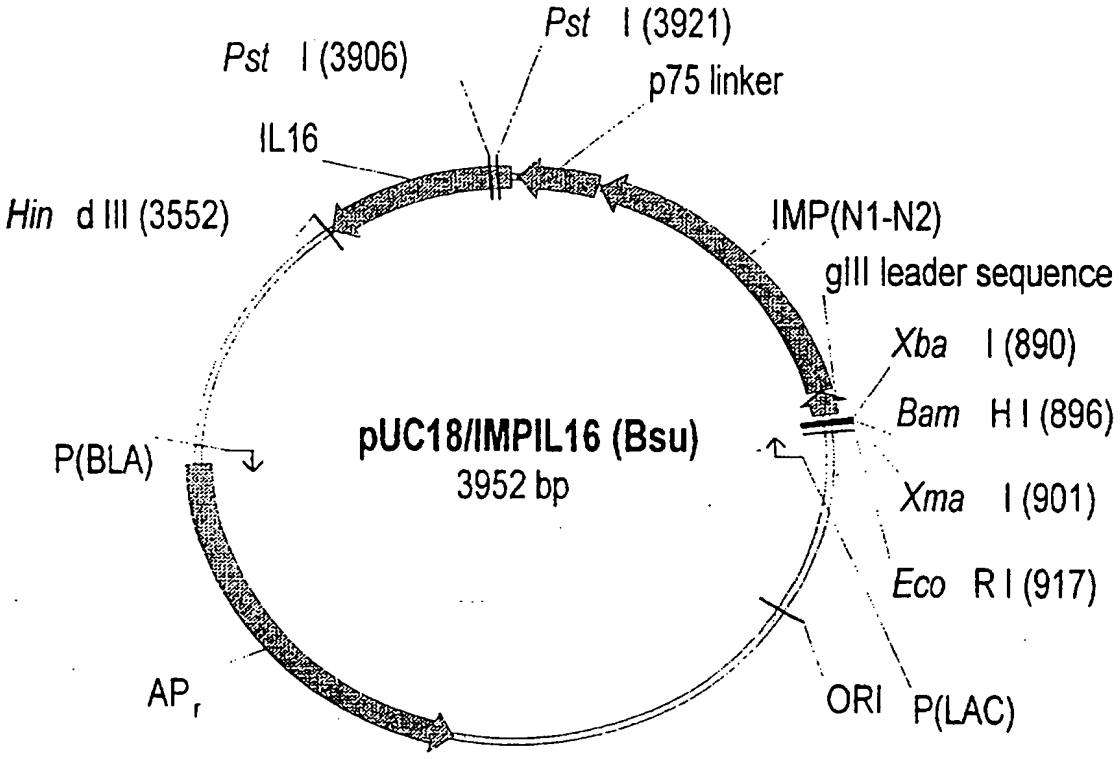


Figure 19: Analysis of selected clones (see Table 3)

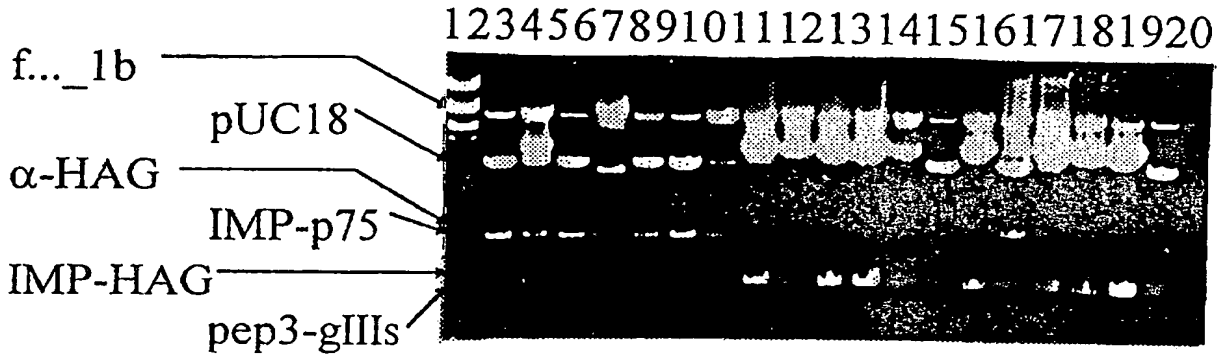


Figure 20: Co-transformation of phagemids, *in vivo* recombination and selection *via* His-tag: general description

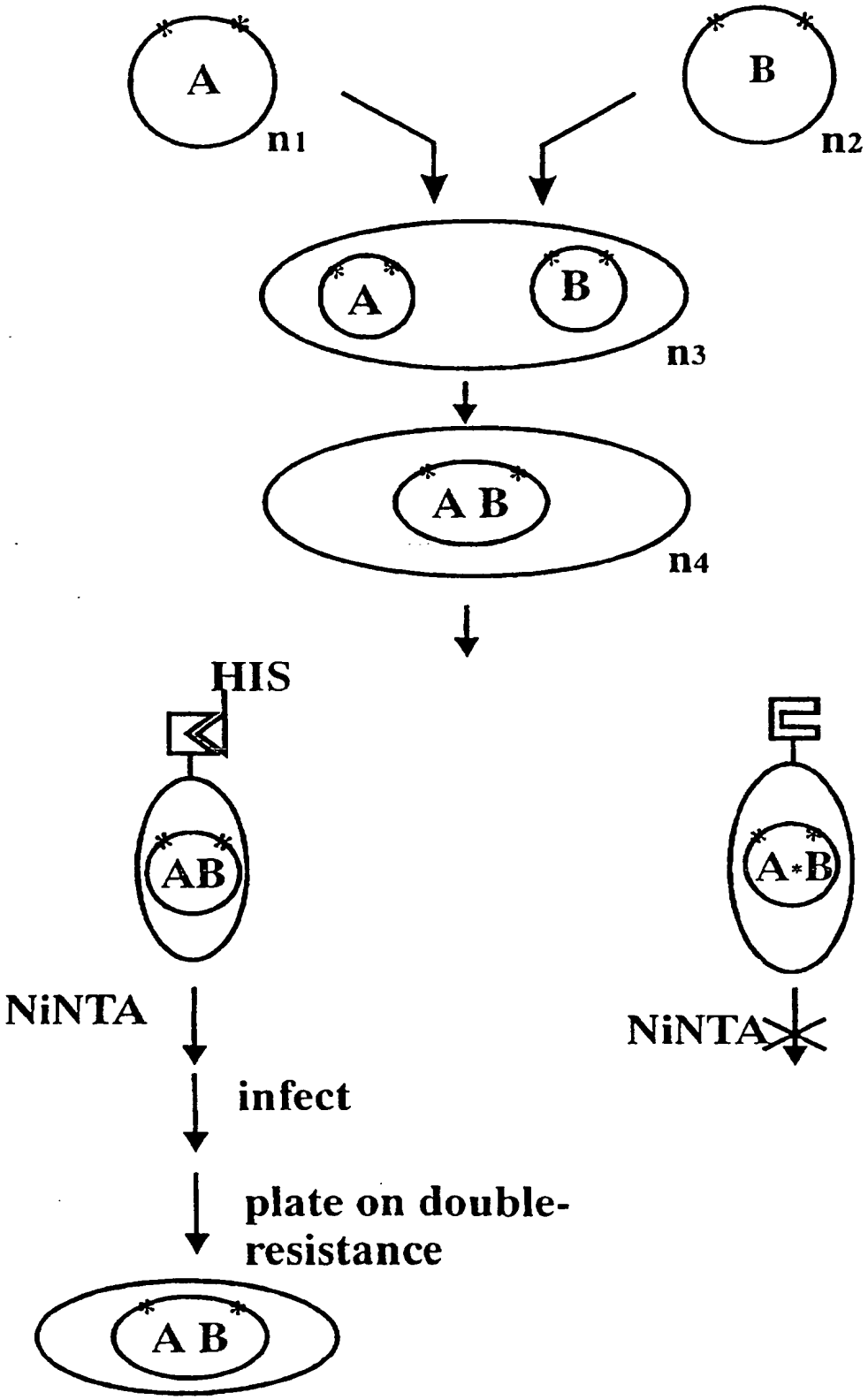


Figure 21: *In vitro* recombination and selection *via* His-tag: general description

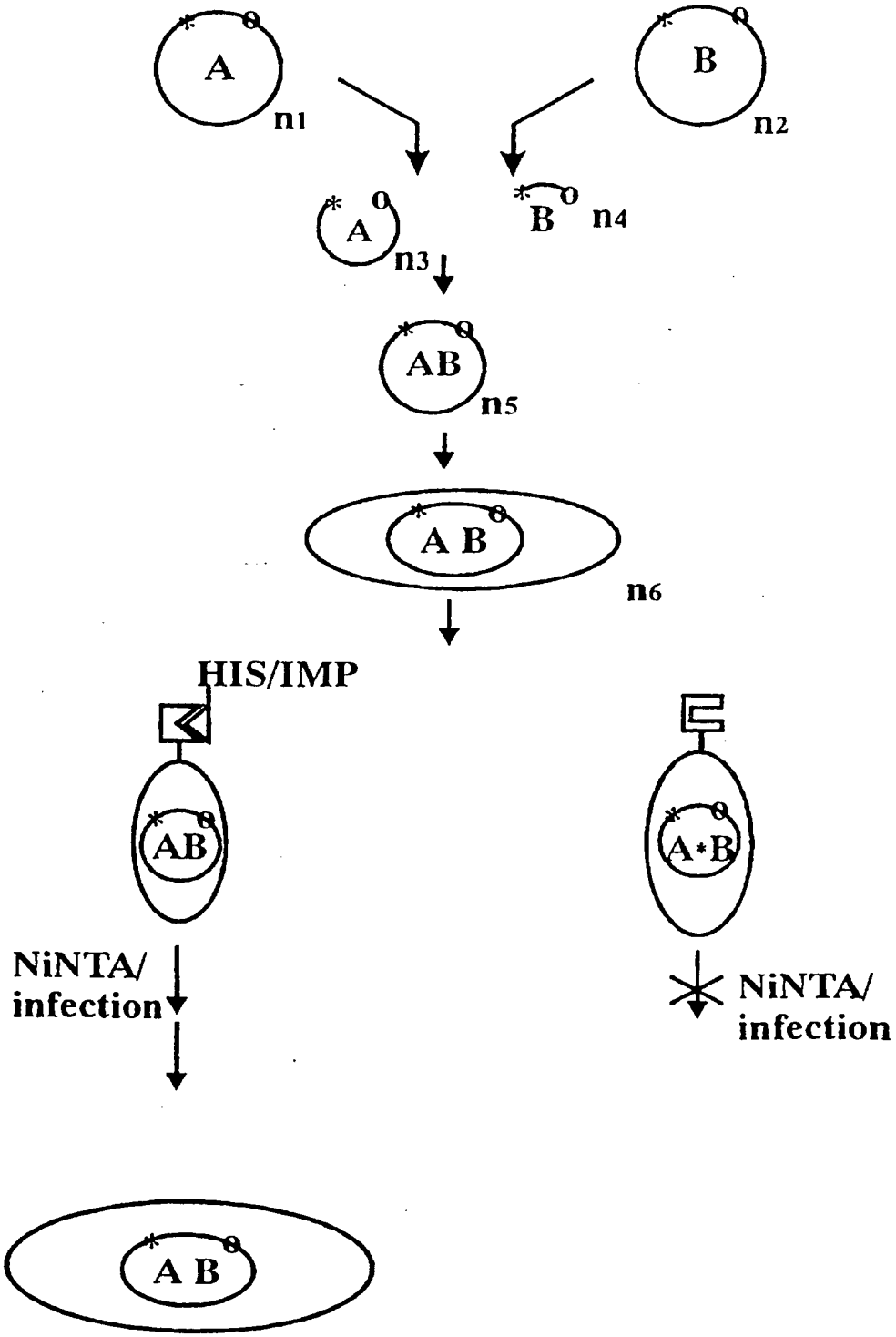


Figure 22: Phage vector fjunhag: functional map

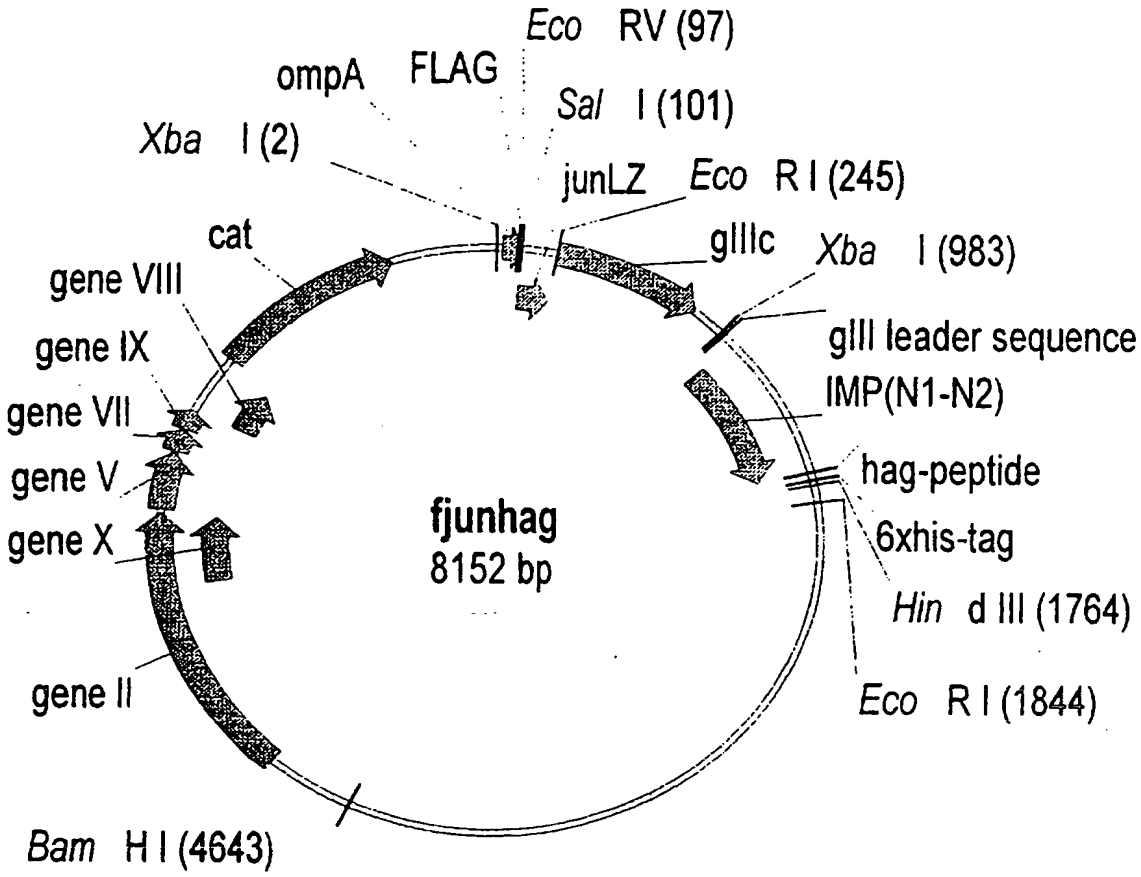


Figure 23: Spatial *in vivo* SIP: general description

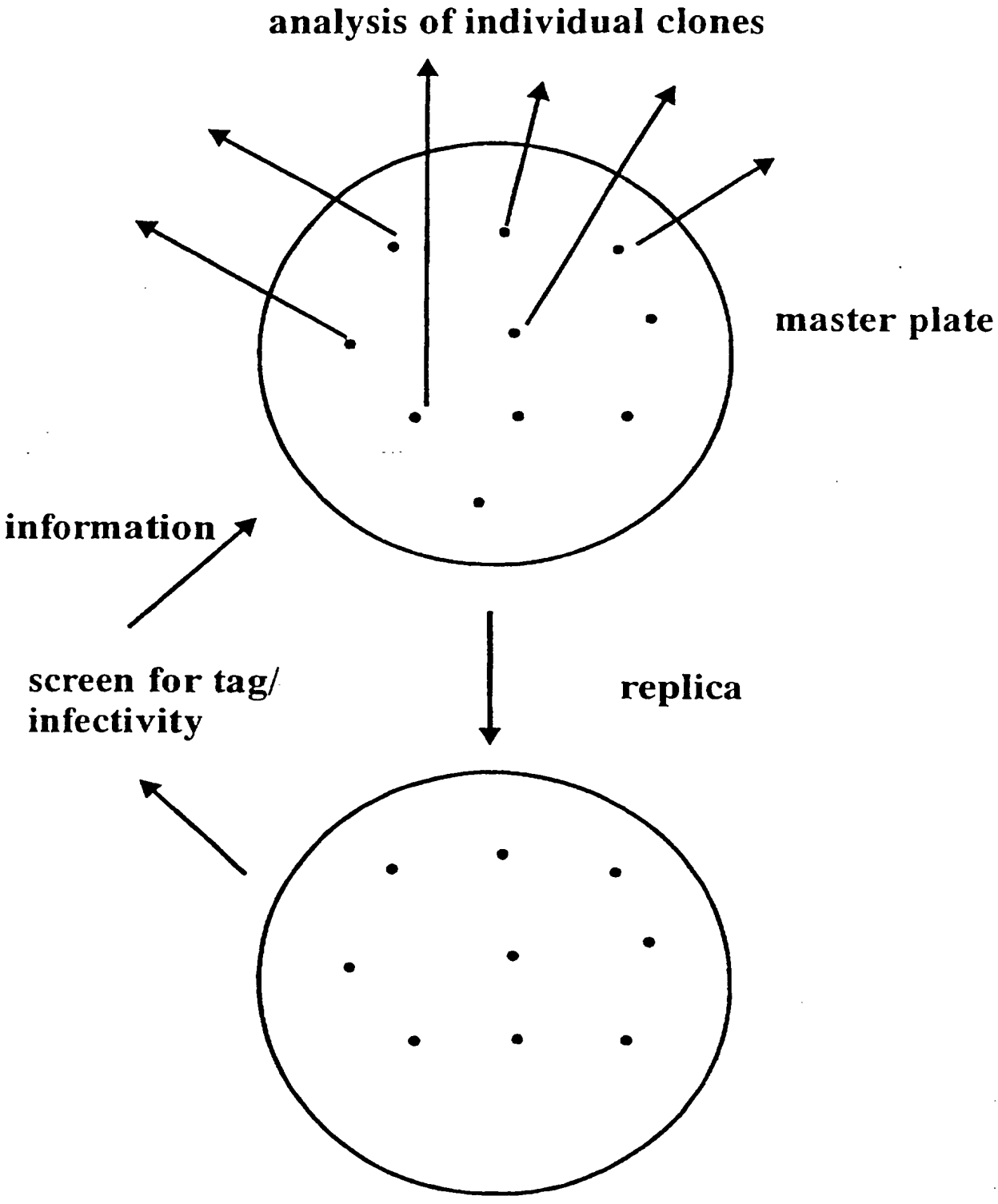


Figure 24: *E. coli* display: general description

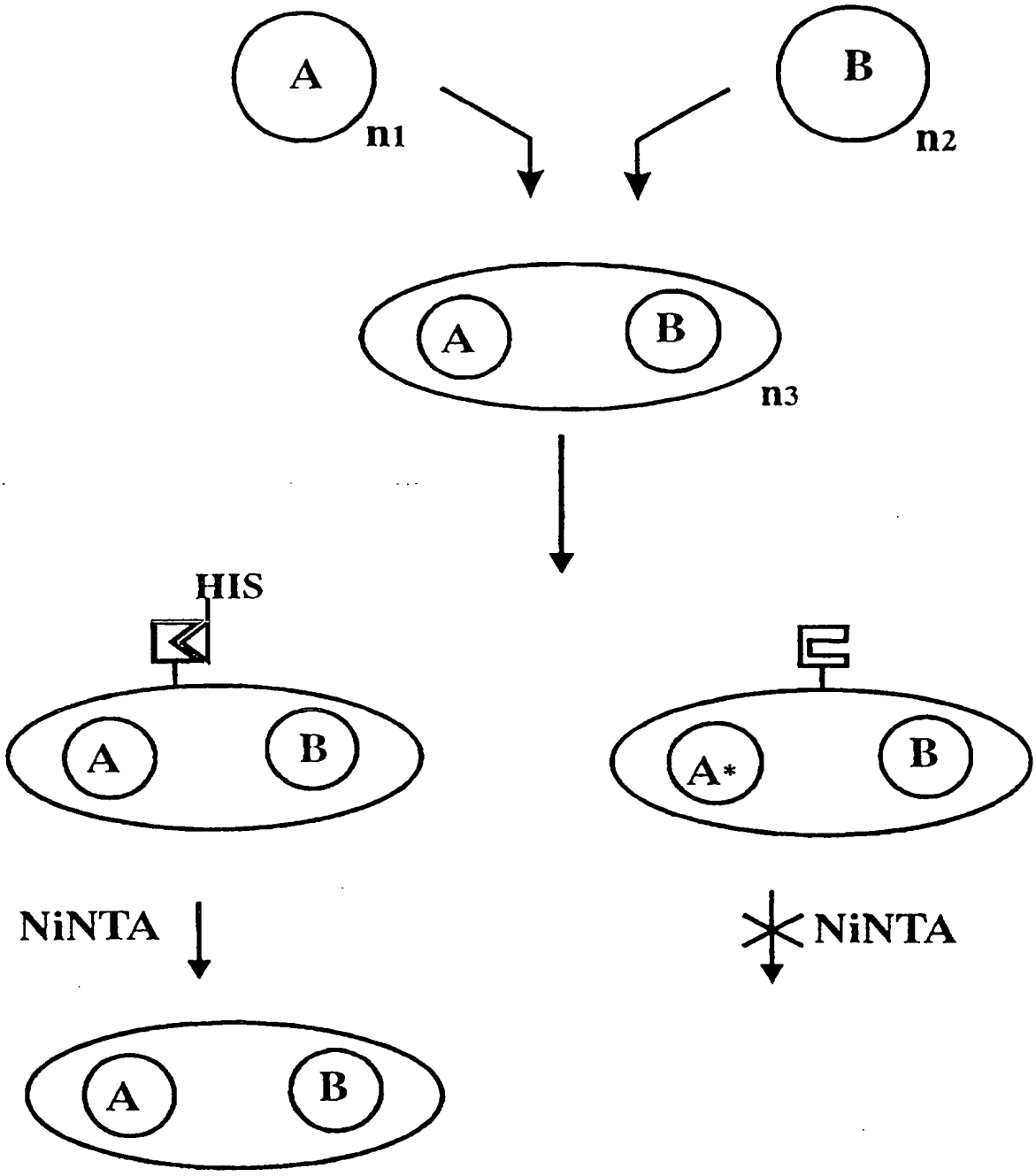


Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence

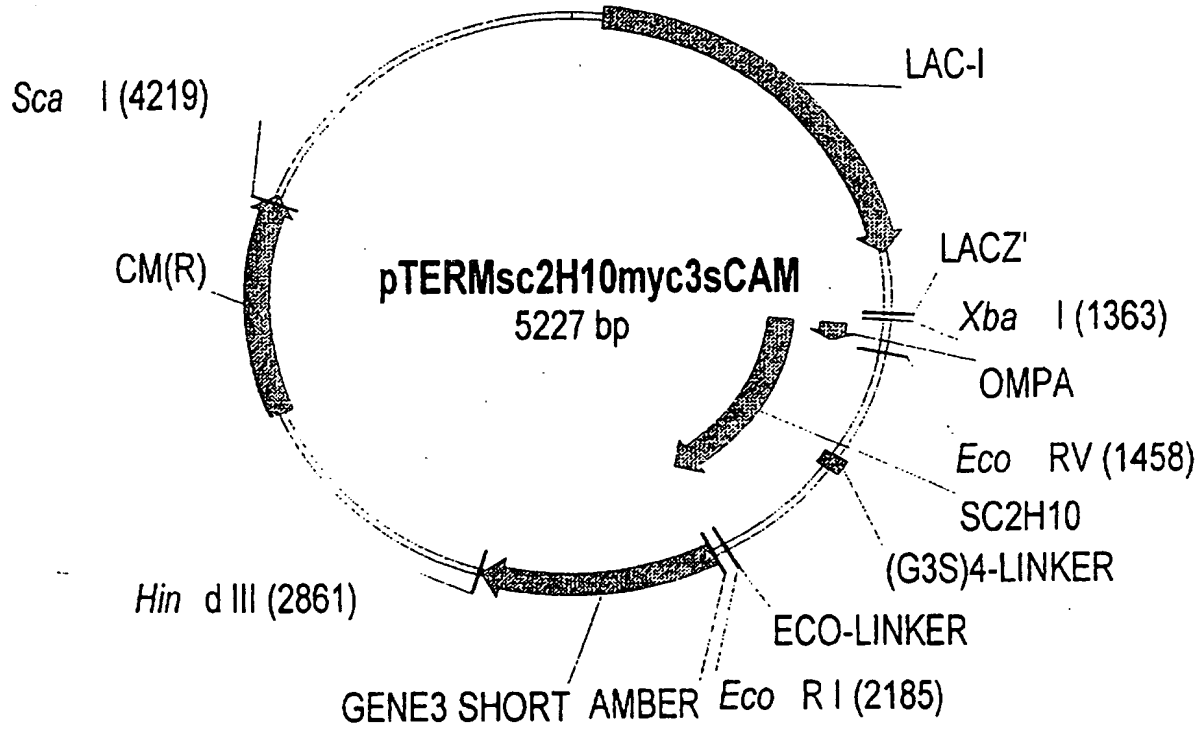


Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
   1  ACCCGACACC ATCGAATGGC GCAAAACCTT TCGCGGTATG GCATGATAGC
      TGGGCTGTGG TAGCTTACCG CGTTTTGGAA AGCGCCATAC CGTACTATCG

  51  GCCCGGAAGA GAGTCAATTC AGGGTGGTGA ATGTGAAACC AGTAACGTTA
      CGGGCCTTCT CTCAGTTAAG TCCCACCACT TACACTTTGG TCATTGCAAT

 101  TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG TTTCCCGCGT
      ATGCTACAGC GTCTCATACG GCCACAGAGA ATAGTCTGGC AAAGGGCGCA

 151  GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAG
      CCACTTGGTC CGGTCGGTGC AAAGACGCTT TTGCGCCCTT TTTCACCTTC

 201  CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG
      GCCGCTACCG CCTCGACTTA ATGTAAGGGT TGGCGCACCG TGTTGTTGAC

 251  GCGGGCAAAC AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT
      CGCCCGTTTG TCAGCAACGA CTAACCGCAA CGGTGGAGGT CAGACCGGGA

 301  GCACGCGCCG TCGCAAATTG TCGCGGCGAT TAAATCTCGC GCCGATCAAC
      CGTGCGCGGC AGCGTTTAAC AGCGCCGCTA ATTTAGAGCG CGGCTAGTTG

 351  TGGGTGCCAG CGTGGTGGTG TCGATGGTAG AACGAAGCGG CGTCGAAGCC
      ACCCACGGTC GCACCACCAC AGCTACCATC TTGCTTCGCC GCAGCTTCGG

 401  TGTAAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA GTGGGCTGAT
      ACATTTCGCC GCCACGTGTT AGAAGAGCGC GTTGCGCAGT CACCCGACTA

 451  CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
      GTAATTGATA GGCGACCTAC TGGTCCTACG GTAACGACAC CTTCGACGGA

 501  GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC
      CGTGATTACA AGGCCGCAAT AAAGAACTAC AGAGACTGGT CTGTGGGTAG

 551  AACAGTATTA TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA
      TTGTCATAAT AAAAGAGGGT ACTTCTGCCA TGCGCTGACC CGCACCTCGT

 601  TCTGGTCGCA TTGGGTCACC AGCAAATCGC GCTGTTAGCG GGCCCATTAA
      AGACCAGCGT AACCCAGTGG TCGTTTAGCG CGACAATCGC CCGGGTAATT

 651  GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG GCTGGCATAA ATATCTCACT
      CAAGACAGAG CCGCGCAGAC GCAGACCGAC CGACCGTATT TATAGAGTGA

 701  CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT GGAGTGCCAT
      GCGTTAGTTT AAGTCGGCTA TCGCCTTGCC CTTCCGCTGA CCTCACGGTA

 751  GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
      CAGGCCAAAA GTTGTTTGGT ACGTTTACGA CTTACTCCCG TAGCAAGGGT

 801  CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC
      GACGCTACGA CCAACGGTTG CTAGTCTACC GCGACCCGCG TTACGCGCGG
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
 851  ATTACCGAGT CCGGGCTGCG CGTTGGTGCG GACATCTCGG TAGTGGGATA
      TAATGGCTCA GGCCCGACGC GCAACCACGC CTGTAGAGCC ATCACCCTAT

 901  CGACGATACC GAAGACAGCT CATGTTATAT CCCGCCGTTA ACCACCATCA
      GCTGCTATGG CTTCTGTCGA GTACAATATA GGGCGGCAAT TGGTGGTAGT

 951  AACAGGATTT TCGCCTGCTG GGGCAAACCA GCGTGGACCG CTTGCTGCAA
      TTGTCCTAAA AGCGGACGAC CCCGTTTGGT CGCACCTGGC GAACGACGTT

1001  CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC CCGTCTCACT
      GAGAGAGTCC CGGTCCGCCA CTTCCCGTTA GTCGACAACG GGCAGAGTGA

1051  GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
      CCACTTTTCT TTTTGGTGGG ACCGCGGGTT ATGCGTTTGG CGGAGAGGGG

1101  GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG
      CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGGCTGAC

1151  GAAAGCGGGC AGTGAGCGGT ACCCGATAAA AGCGGCTTCC TGACAGGAGG
      CTTTCGCCCG TCACTCGCCA TGGGCTATTT TCGCCGAAGG ACTGTCCTCC

1201  CCGTTTTGTT TTGCAGCCCA CCTCAACGCA ATTAATGTGA GTTAGCTCAC
      GGCAAAACAA AACGTCGGGT GGAGTTGCGT TAATTACACT CAATCGAGTG

1251  TCATTAGGCA CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT
      AGTAATCCGT GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA

1301  GTGGAATTGT GAGCGGATAA CAATTTCACA CAGGAAACAG CTATGACCAT
      CACCTTAACA CTCGCCTATT GTTAAAGTGT GTCCTTTGTC GATACTGGTA

                  XbaI
                  ~~~~~~
1351  GATTACGAAT TTCTAGATAA CGAGGGCAAA AAATGAAAAA GACAGCTATC
      CTAATGCTTA AAGATCTATT GCTCCCGTTT TTTACTTTTT CTGTCGATAG

1401  GCGATTGCAG TGGCACTGGC TGGTTTCGCT ACCGTAGCGC AGGCCGACTA
      CGCTAACGTC ACCGTGACCG ACCAAAGCGA TGGCATCGCG TCCGGCTGAT

        EcoRV
        ~~~~~~
1451  CAAAGATATC GTGATGACCC AGTCTCCAGC AATCATGTCT ACATCTCTAG
      GTTTCTATAG CACTACTGGG TCAGAGGTCG TTAGTACAGA TGTAGAGATC

1501  GGGAACGGGT CACCATGACC TGCACTGCCA GTTCAAGTGT AAGTTCCTCT
      CCCTTGCCCA GTGGTACTGG ACGTGACGGT CAAGTTCACA TTCAAGGAGA

1551  TACTTACACT GGTACCAGCA GAAGCCAGGA TCCTCCCCCA AACTCTGGAT
      ATGAATGTGA CCATGGTCGT CTTCGGTCCT AGGAGGGGGT TTGAGACCTA

1601  TTATAGCACA TCCAACCTGG CTTCTGGAGT CCCAACTCGC TTCAGTGGCA
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
        AATATCGTGT AGGTTGGACC GAAGACCTCA GGGTTGAGCG AAGTCACCGT

1651    GTGGGTCTGG GACCTCTTAC TCTCTCACAA TCAGCACCAT GGCGGCTGAG
        CACCCAGACC CTGGAGAATG AGAGAGTGTT AGTCGTGGTA CCGCCGACTC

1701    GATGCTGCCA CTTATTACTG CCACCAGTAT CATCGTTTCC CACCCACGTT
        CTACGACGGT GAATAATGAC GGTGGTCATA GTAGCAAAGG GTGGGTGCAA

1751    CGGAGGGGGG ACCAAGCTGG AAATAAAACG GGCTGGTGGT GGTGGTTCTG
        GCCTCCCCCC TGGTTCGACC TTTATTTTGC CCGACCACCA CCACCAAGAC

1801    GCGGCGGCGG CTCCGGTGGT GGTGGTTCTG AAGTTAAACT GGTCGAGTCT
        CGCCGCCGCC GAGGCCACCA CCACCAAGAC TTCAATTTGA CCAGCTCAGA

1851    GGAGGAGGCT TGGTGCAACC TGGAGGATCC ATGAAACTCT CCTGTGTTGC
        CCTCCTCCGA ACCACGTTGG ACCTCCTAGG TACTTTGAGA GGACACAACG

1901    CTCTGGAATC ACTTTCAGTA ATTACCGGAT GAACTGGGTC CGCCAGTCTC
        GAGACCTTAG TGAAAGTCAT TAATGGCCTA CTTGACCCAG GCGGTCAGAG

1951    CAGAGAAGGG GCTTGAGTGG GTTGCTGAAA TTAGATTGAA ATCTAATAAT
        GTCTCTTCCC CGAACTCACC CAACGACTTT AATCTAACTT TAGATTATTA

2001    TATGCAACAC ATTATGCGGA GTCTGTGAAA GGGAGGTTCA CCATCTCAAG
        ATACGTTGTG TAATACGCCT CAGACACTTT CCCTCCAAGT GGTAGAGTTC

2051    AGATGATTCC AAAAGTAGTG TCTACCTGCA AATGAACAAC TTAAGAGCTG
        TCTACTAAGG TTTTCATCAC AGATGGACGT TTACTTGTTG AATTCTCGAC

2101    AAGACACTGG CATTTATTAC TGTAGAGGGG TTTCATATAC TATAGACTAC
        TTCTGTGACC GTAAATAATG ACATCTCCCC AAAGTATATG ATATCTGATG

                                                  EcoRI
                                                  ~~~~~~
2151    TGGGGTCAAG GAACCTCAGT CACAGTCTCC TCAGAATTCG AGCAGAAGCT
        ACCCCAGTTC CTTGGAGTCA GTGTCAGAGG AGTCTTAAGC TCGTCTTCGA

2201    GATCTCTGAG GAAGACCTGT AGGCATGCTT ATTTGTTTGT GAATATCAAG
        CTAGAGACTC CTTCTGGACA TCCGTACGAA TAAACAAACA CTTATAGTTC

2251    GCCAATCGTC TGACCTGCCT CAACCTCCTG TCAATGCTGG CGGCGGCTCT
        CGGTTAGCAG ACTGGACGGA GTTGGAGGAC AGTTACGACC GCCGCCGAGA

2301    GGTGGTGGTT CTGGTGGCGG CTCTGAGGGT GGTGGCTCTG AGGGTGGCGG
        CCACCACCAA GACCACCGCC GAGACTCCCA CCACCGAGAC TCCCACCGCC

2351    TTCTGAGGGT GGCGGCTCTG AGGGAGGCGG TTCCGGTGGT GGCTCTGGTT
        AAGACTCCCA CCGCCGAGAC TCCCTCCGCC AAGGCCACCA CCGAGACCAA

2401    CCGGTGATTT TGATTATGAA AAGATGGCAA ACGCTAATAA GGGGGCTATG
        GGCCACTAAA ACTAATACTT TTCTACCGTT TGCGATTATT CCCCCGATAC
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
2451  ACCGAAAATG CCGATGAAAA CGCGCTACAG TCTGACGCTA AAGGCAAACT
      TGGCTTTTAC GGCTACTTTT GCGCGATGTC AGACTGCGAT TTCCGTTTGA

2501  TGATTCTGTC GCTACTGATT ACGGTGCTGC TATCGATGGT TTCATTGGTG
      ACTAAGACAG CGATGACTAA TGCCACGACG ATAGCTACCA AAGTAACCAC

2551  ACGTTTCCGG CCTTGCTAAT GGTAATGGTG CTACTGGTGA TTTTGCTGGC
      TGCAAAGGCC GGAACGATTA CCATTACCAC GATGACCACT AAAACGACCG

2601  TCTAATTCCC AAATGGCTCA AGTCGGTGAC GGTGATAATT CACCTTTAAT
      AGATTAAGGG TTTACCGAGT TCAGCCACTG CCACTATTAA GTGGAAATTA

2651  GAATAATTTC CGTCAATATT TACCTTCCCT CCCTCAATCG GTTGAATGTC
      CTTATTAAAG GCAGTTATAA ATGGAAGGGA GGGAGTTAGC CAACTTACAG

2701  GCCCTTTTGT CTTTGGCGCT GGTAAACCAT ATGAATTTTC TATTGATTGT
      CGGGAAAACA GAAACCGCGA CCATTTGGTA TACTTAAAAG ATAACTAACA

2751  GACAAAATAA ACTTATTCCG TGGTGTCTTT GCGTTTCTTT TATATGTTGC
      CTGTTTTATT TGAATAAGGC ACCACAGAAA CGCAAAGAAA ATATACAACG

2801  CACCTTTATG TATGTATTTT CTACGTTTGC TAACATACTG CGTAATAAGG
      GTGGAAATAC ATACATAAAA GATGCAAACG ATTGTATGAC GCATTATTCC

                HindIII
                ~~~~~~~
2851  AGTCTTGATA AGCTTGACCT GTGAAGTGAA AAATGGCGCA CATTGTGCGA
      TCAGAACTAT TCGAACTGGA CACTTCACTT TTTACCGCGT GTAACACGCT

2901  CATTTTTTTT GTCTGCCGTT TACCGCTACT GCGTCACGGA TCCCCACGCG
      GTAAAAAAAA CAGACGGCAA ATGGCGATGA CGCAGTGCCT AGGGGTGCGC

2951  CCCTGTAGCG GCGCATTAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT
      GGGACATCGC CGCGTAATTC GCGCCGCCCA CACCACCAAT GCGCGTCGCA

3001  GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTTCTTCC
      CTGGCGATGT GAACGGTCGC GGGATCGCGG GCGAGGAAAG CGAAAGAAGG

3051  CTTCCTTTCT CGCCACGTTC GCCGGCTTTC CCCGTCAAGC TCTAAATCGG
      GAAGGAAAGA GCGGTGCAAG CGGCCGAAAG GGGCAGTTCG AGATTTAGCC

3101  GGCATCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA
      CCGTAGGGAA ATCCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGGGTT

3151  AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA
      TTTTGAACTA ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT

3201  CGGTTTTTCG CCCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGACTC
      GCCAAAAAGC GGGAAACTGC AACCTCAGGT GCAAGAAATT ATCACCTGAG
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
3251  TTGTTCCAAA CTGGAACAAC ACTCAACCCT ATCTCGGTCT ATTCTTTTGA
      AACAAGGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACT

3301  TTTATAAGGG ATTTTGCCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA
      AAATATTCCC TAAAACGGCT AAAGCCGGAT AACCAATTTT TTACTCGACT

3351  TTTAACAAAA ATTTAACGCG AATTTTAACA AAATATTAAC GTTTACAATT
      AAATTGTTTT TAAATTGCGC TTAAAATTGT TTTATAATTG CAAATGTTAA

3401  TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GGAACCCCTA TTTGTTTATT
      AGTCCACCGT GAAAAGCCCC TTTACACGCG CCTTGGGGAT AAACAAATAA

3451  TTTCTAAATA CATTCAAATA TGTATCCGCT CATGTCGAGA CGTTGGGTGA
      AAAGATTTAT GTAAGTTTAT ACATAGGCGA GTACAGCTCT GCAACCCACT

3501  GGTTCCAACT TTCACCATAA TGAAATAAGA TCACTACCGG GCGTATTTTT
      CCAAGGTTGA AAGTGGTATT ACTTTATTCT AGTGATGGCC CGCATAAAAA

3551  TGAGTTATCG AGATTTTCAG GAGCTAAGGA AGCTAAAATG GAGAAAAAAA
      ACTCAATAGC TCTAAAAGTC CTCGATTCCT TCGATTTTAC CTCTTTTTTT

3601  TCACTGGATA TACCACCGTT GATATATCCC AATGGCATCG TAAAGAACAT
      AGTGACCTAT ATGGTGGCAA CTATATAGGG TTACCGTAGC ATTTCTTGTA

3651  TTTGAGGCAT TTCAGTCAGT TGCTCAATGT ACCTATAACC AGACCGTTCA
      AAACTCCGTA AAGTCAGTCA ACGAGTTACA TGGATATTGG TCTGGCAAGT

3701  GCTGGATATT ACGGCCTTTT TAAAGACCGT AAAGAAAAAT AAGCACAAGT
      CGACCTATAA TGCCGGAAAA ATTTCTGGCA TTTCTTTTTA TTCGTGTTCA

3751  TTTATCCGGC CTTTATTCAC ATTCTTGCCC GCCTGATGAA TGCTCATCCG
      AAATAGGCCG GAAATAAGTG TAAGAACGGG CGGACTACTT ACGAGTAGGC

3801  GAGTTCCGTA TGGCAATGAA AGACGGTGAG CTGGTGATAT GGGATAGTGT
      CTCAAGGCAT ACCGTTACTT TCTGCCACTC GACCACTATA CCCTATCACA

3851  TCACCCTTGT TACACCGTTT TCCATGAGCA AACTGAAACG TTTTCATCGC
      AGTGGGAACA ATGTGGCAAA AGGTACTCGT TTGACTTTGC AAAAGTAGCG

3901  TCTGGAGTGA ATACCACGAC GATTTCCGGC AGTTTCTACA CATATATTCG
      AGACCTCACT TATGGTGCTG CTAAAGGCCG TCAAAGATGT GTATATAAGC

3951  CAAGATGTGG CGTGTTACGG TGAAAACCTG GCCTATTTCC CTAAAGGGTT
      GTTCTACACC GCACAATGCC ACTTTTGGAC CGGATAAAGG GATTTCCCAA

4001  TATTGAGAAT ATGTTTTTCG TCTCAGCCAA TCCCTGGGTG AGTTTCACCA
      ATAACTCTTA TACAAAAAGC AGAGTCGGTT AGGGACCCAC TCAAAGTGGT

4051  GTTTTGATTT AAACGTGGCC AATATGGACA ACTTCTTCGC CCCCGTTTTC
      CAAAACTAAA TTTGCACCGG TTATACCTGT TGAAGAAGCG GGGGCAAAAG
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
4101  ACCATGGGCA AATATTATAC GCAAGGCGAC AAGGTGCTGA TGCCGCTGGC
      TGGTACCCGT TTATAATATG CGTTCCGCTG TTCCACGACT ACGGCGACCG

4151  GATTCAGGTT CATCATGCCG TCTGTGATGG CTTCCATGTC GGCAGAATGC
      CTAAGTCCAA GTAGTACGGC AGACACTACC GAAGGTACAG CCGTCTTACG

                          ScaI
                          ~~~~~~~
4201  TTAATGAATT ACAACAGTAC TGCGATGAGT GGCAGGGCGG GGCGTAATTT
      AATTACTTAA TGTTGTCATG ACGCTACTCA CCGTCCCGCC CCGCATTAAA

4251  TTTTAAGGCA GTTATTGGTG CCCTTAAACG CCTGGTGCTA CGCCTGAATA
      AAAATTCCGT CAATAACCAC GGGAATTTGC GGACCACGAT GCGGACTTAT

4301  AGTGATAATA AGCGGATGAA TGGCAGAAAT TCGAAAGCAA ATTCGACCCG
      TCACTATTAT TCGCCTACTT ACCGTCTTTA AGCTTTCGTT TAAGCTGGGC

4351  GTCGTCGGTT CAGGGCAGGG TCGTTAAATA GCCGCTTATG TCTATTGCTG
      CAGCAGCCAA GTCCCGTCCC AGCAATTTAT CGGCGAATAC AGATAACGAC

4401  GTTTACCGGT TTATTGACTA CCGGAAGCAG TGTGACCGTG TGCTTCTCAA
      CAAATGGCCA AATAACTGAT GGCCTTCGTC ACACTGGCAC ACGAAGAGTT

4451  ATGCCTGAGG CCAGTTTGCT CAGGCTCTCC CCGTGGAGGT AATAATTGCT
      TACGGACTCC GGTCAAACGA GTCCGAGAGG GGCACCTCCA TTATTAACGA

4501  CGACATGACC AAAATCCCTT AACGTGAGTT TTCGTTCCAC TGAGCGTCAG
      GCTGTACTGG TTTTAGGGAA TTGCACTCAA AAGCAAGGTG ACTCGCAGTC

4551  ACCCCGTAGA AAAGATCAAA GGATCTTCTT GAGATCCTTT TTTTCTGCGC
      TGGGGCATCT TTTCTAGTTT CCTAGAAGAA CTCTAGGAAA AAAAGACGCG

4601  GTAATCTGCT GCTTGCAAAC AAAAAAACCA CCGCTACCAG CGGTGGTTTG
      CATTAGACGA CGAACGTTTG TTTTTTTGGT GGCGATGGTC GCCACCAAAC

4651  TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA
      AAACGGCCTA GTTCTCGATG GTTGAGAAAA AGGCTTCCAT TGACCGAAGT

4701  GCAGAGCGCA GATACCAAAT ACTGTCCTTC TAGTGTAGCC GTAGTTAGGC
      CGTCTCGCGT CTATGGTTTA TGACAGGAAG ATCACATCGG CATCAATCCG

4751  CACCACTTCA AGAACTCTGT AGCACCGCCT ACATACCTCG CTCTGCTAAT
      GTGGTGAAGT TCTTGAGACA TCGTGGCGGA TGTATGGAGC GAGACGATTA

4801  CCTGTTACCA GTGGCTGCTG CCAGTGGCGA TAAGTCGTGT CTTACCGGGT
      GGACAATGGT CACCGACGAC GGTCACCGCT ATTCAGCACA GAATGGCCCA

4851  TGGACTCAAG ACGATAGTTA CCGGATAAGG CGCAGCGGTC GGGCTGAACG
      ACCTGAGTTC TGCTATCAAT GGCCTATTCC GCGTCGCCAG CCCGACTTGC

4901  GGGGGTTCGT GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGAACT
```

Figure 25: pTERMsc2H10myc3sCAM: functional map and sequence (continued)

```
        CCCCCAAGCA CGTGTGTCGG GTCGAACCTC GCTTGCTGGA TGTGGCTTGA

4951    GAGATACCTA CAGCGTGAGC TATGAGAAAG CGCCACGCTT CCCGAAGGGA
        CTCTATGGAT GTCGCACTCG ATACTCTTTC GCGGTGCGAA GGGCTTCCCT

5001    GAAAGGCGGA CAGGTATCCG GTAAGCGGCA GGGTCGGAAC AGGAGAGCGC
        CTTTCCGCCT GTCCATAGGC CATTCGCCGT CCCAGCCTTG TCCTCTCGCG

5051    ACGAGGGAGC TTCCAGGGGG AAACGCCTGG TATCTTTATA GTCCTGTCGG
        TGCTCCCTCG AAGGTCCCCC TTTGCGGACC ATAGAAATAT CAGGACAGCC

5101    GTTTCGCCAC CTCTGACTTG AGCGTCGATT TTTGTGATGC TCGTCAGGGG
        CAAAGCGGTG GAGACTGAAC TCGCAGCTAA AAACACTACG AGCAGTCCCC

5151    GGCGGAGCCT ATGGAAAAAC GCCAGCAACG CGGCCTTTTT ACGGTTCCTG
        CCGCCTCGGA TACCTTTTTG CGGTCGTTGC GCCGGAAAAA TGCCAAGGAC

5201    GCCTTTTGCT GGCCTTTTGC TCACATG
        CGGAAAACGA CCGGAAAACG AGTGTAC
```

Table 1: Phagemids Constructed for Experiments 2 and 3

Name	FLAG	His6	gIII	Size (bp)	Insert	REN1	REN2	Resistance
pING1-A1	-	+	-	3783	His	EcoRV	SmaI	Ap
pING1-A2	-	-	-	3795	Strep-tag	EcoRV	SmaI	Ap
pING3-A1	+	+	-	3792	His	EcoRV	SmaI	Ap
pING3-A2	+	-	-	3804	Strep-tag	EcoRV	SmaI	Ap
pONG3-A	+	-	+	4278	-	EcoRV	SmaI	Ap
pYANG3-A	+	-	+	4404	Jun	EcoRV	EcoRI	Ap
pYANG3-Ape2	+	-	+		pep2	XbaI	HindIII	Ap
pYANG3-Ape3	+	-	+		pep3	XbaI	HindIII	Ap
pYANG3-Ape10	+	-	+		pep10	XbaI	HindIII	Ap
pING1-C1	-	+	-	3853	His	EcoRV	SmaI	Cm
pING1-C2	-	-	-	3865	Strep-tag	EcoRV	SmaI	Cm
pING3-C1	+	+	-	3862	His	EcoRV	SmaI	Cm
pING3-C2	+	-	-	3874	Strep-tag	EcoRV	SmaI	Cm
pYING3-C1	+	+	-	3994	Fos	EcoRV	EcoRI	Cm
pYING3-C2	+	+	-	4315	p75	EcoRV	EcoRI	Cm
pYING3-C3	+	+	-	4240	IL-16	EcoRV	EcoRI	Cm

Table 2: Results of Experiment 2 (see Figure 7)

Table 2a: Combination of phagemids present in initial library (α)

	Combination	Clone(s)
1.	pYING1-C2 + pYANG3-ApeX	6
2.	pYING1-C1 + pYANG3-A	1
3.	pYING1-C1 + pYANG3-ApeX	1
4.	pYING1-C2 + pYANG3-A	1
5.	pYING1-C2 + ?	1

Table 2b: Combination of phagemids present after selection (β)

	Combination	Clone(s)
1.	pYING1-C2 + pYANG3-ApeX	1
2.	pYING1-C1 + pYANG3-A	9

Table 3: Results of Experiment 4 (see Figure 19)

Table 3a: Identification of phage/plasmid present in individual clones

Combination	Clone(s)
fhag1A + pUC19/IMPhag	#9
fpep3_1b + pUC18/IMP-p75	#1,#3,#5,#6,#7,#13,#15,#19
fpep3_1b + pUC19/IMPhag	#14
unusual DNA	#2,#4,#8,#10,#11,#12,#16,#17,#18

Table 3b: Test for infectivity of individual clones

Clone #	Titer (transducing units/ml)
1	2 x 10E4
2	31
3	1 x 10E5
4	1 x 10E5
5	1 x 10E5
6	2 x 10E3
7	1 x 10E4
8	1 x 10E5
9	1 x 10E6
10	1 x 10E4
11	1 x 10E3
12	1 x 10E4
13	3 x 10E3
14	< 10
15	5 x 10E4
16	1 x 10E4
17	5 x 10E2
18	1 x 10E4
19	1 x 10E5

INTERNATIONAL SEARCH REPORT

Intern al Application No
PCT/EP 97/00931

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/10 C12N15/12 C12N15/24 C12N15/62 C12N15/70

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 93 19172 A (CAMBRIDGE ANTIBODY TECH ;MEDICAL RES COUNCIL (GB)) 30 September 1993 see the whole document ---	1-7,11, 12, 19-27, 29,30,32
A	EP 0 614 989 A (MORPHOSYS GES FUER PROTEINOPTI) 14 September 1994 cited in the application see the whole document ---	1-34
A	DE 41 22 598 C (DKFZ) 30 July 1992 see the whole document --- -/--	1-34

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
19 June 1997	03.07.97

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	HORNIG, H

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interr. al Application No
PCT/EP 97/00931

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GENE, vol. 137, 1993, ELSEVIER SCIENCE PUBLISHERS,B.V.,AMSTERDAM,NL;, pages 69-75, XP002033343 R. CAMERI AND M. SUTER: "Display of biological active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to genetic information responsible for their production" cited in the application see the whole document ---	1-34
A	NUCLEIC ACIDS RESEARCH, vol. 22, no. 25, 1994, IRL PRESS LIMITED,OXFORD,ENGLAND, pages 5761-5762, XP002033344 K. GRAMATIKOFF ET AL.: "Direct interaction rescue, a novel filamentous phage technique to study protein - protein interactions" cited in the application see the whole document ---	1-34
A	NATURE BIOTECHNOLOGY, vol. 13, April 1995, NATURE PUBL. CO.,NEW YORK, US, pages 373-377, XP002033345 D. NERI ET AL.: "Calmodulin as a versatile tag for antibody fragments" see the whole document ---	1-34
A	NATURE BIOTECHNOLOGY, vol. 13, April 1995, NATURE PUBL. CO.,NEW YORK, US, pages 366-372, XP002033346 Z. LU ET AL.: "Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: A system designed for exploring protein-protein interactions" cited in the application see the whole document ---	1-34
A	FEBS LETTERS, vol. 377, no. 2, 18 December 1995, ELSEVIER, AMSTERDAM, NL, pages 227-231, XP002033347 C. KREBBER ET AL.: "Co-selection of cognate antibody-antigen pairs by selectively-infective phage" cited in the application see the whole document --- -/--	1-34

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte ɔnal Application No
PCT/EP 97/00931

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 97 10330 A (BIOINVENT INT AB ;KIDDLE SIMON JOHN (GB); BORREBAECK CARL (SE); MA) 20 March 1997 see the whole document	1,2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

...ormation on patent family members

Intern ...l Application No
PCT/EP 97/00931

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9319172 A	30-09-93	AU 665190 B	21-12-95
		AU 1693892 A	30-12-92
		AU 665025 B	14-12-95
		AU 2593392 A	27-04-93
		AU 665221 B	21-12-95
		AU 3089092 A	28-06-93
		AU 673515 B	14-11-96
		AU 3763893 A	21-10-93
		CA 2109602 A	26-11-92
		CA 2119930 A	01-04-93
		CA 2124460 A	10-06-93
		CA 2131151 A	30-09-94
		EP 0585287 A	09-03-94
		EP 0605522 A	13-07-94
		EP 0616640 A	28-09-94
		EP 0656941 A	14-06-95
		WO 9220791 A	26-11-92
		WO 9306213 A	01-04-93
		WO 9311236 A	10-06-93
		JP 6510671 T	01-12-94
		JP 6508511 T	29-09-94
		JP 7502167 T	09-03-95
		JP 7505055 T	08-06-95
		US 5565332 A	15-10-96
EP 0614989 A	14-09-94	CA 2115811 A	18-08-94
		DE 614989 T	28-09-95
		JP 7039381 A	10-02-95
		US 5514548 A	07-05-96
DE 4122598 C	30-07-92	WO 9301287 A	21-01-93
		EP 0547200 A	23-06-93
		JP 6501395 T	17-02-94
		US 5591604 A	07-01-97
WO 9710330 A	20-03-97	AU 6939196 A	01-04-97

ELSEVIER

Journal of Immunological Methods 231 (1999) 93-104

JOURNAL OF IMMUNOLOGICAL METHODS

www.elsevier.nl/locate/jim

Selectively infective phage (SIP) technology: scope and limitations

Sabine Jung, Katja M. Arndt, Kristian M. Müller, Andreas Plückthun *

Biochemisches Institut, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland

Abstract

We review here the selectively infective phage (SIP) technology, a powerful tool for the rapid selection of protein–ligand and peptide–ligand pairs with very high affinities. SIP is highly suitable for discriminating between molecules with subtle stability and folding differences. We discuss the preferred types of applications for this technology and some pitfalls inherent in the in vivo SIP method that have become apparent in its application with highly randomized libraries, as well as some precautions that should be taken in successfully applying this technology. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Phage display; Affinity maturation; Disulfide bonds; ScFv fragments

1. Introduction

1.1. The principle of SIP

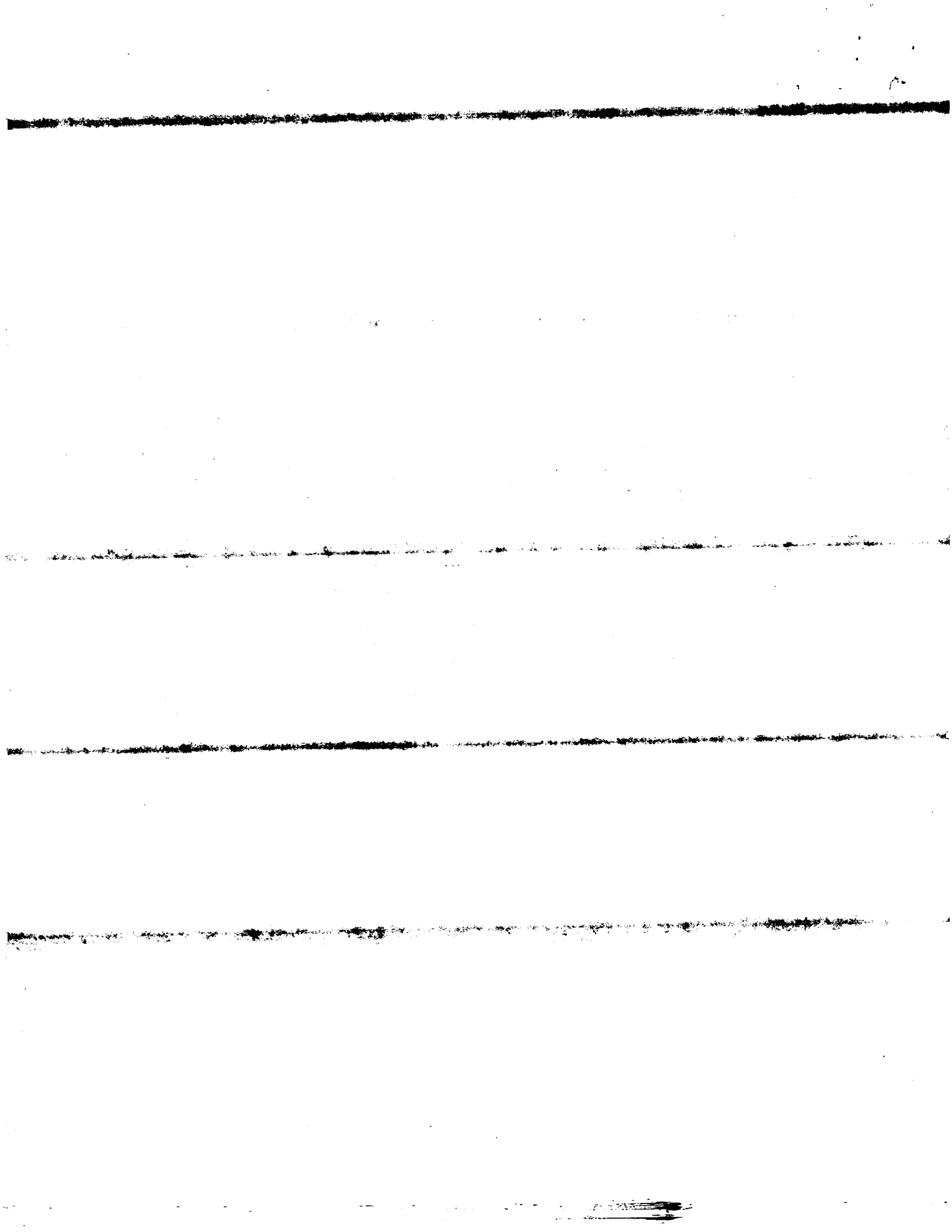
The selectively infective phage (SIP) technology was developed for selecting interacting protein–ligand pairs (Dueñas and Borrebaeck, 1994; Gramatikoff et al., 1994; Krebber et al., 1995). It has also been called selection and amplification of phage (SAP) (Dueñas and Borrebaeck, 1994) or direct interaction rescue (DIRE) (Gramatikoff et al., 1994). While SIP is related to phage display, it has the advantage of directly coupling the productive protein–ligand interaction with phage infectivity and amplification, without the need of an elution step from a solid matrix (Fig. 1).

SIP exploits the modular structure of the gene-3-protein (g3p), which consists of three domains, N1, N2 and CT, which are connected by glycine-rich linkers and possess different functions for the phage life cycle (Fig. 1) (Armstrong et al., 1981; Stengele et al., 1990). The g3p is present most likely in five copies on the phage, reflecting the five-fold symmetry of the phage coat and the pilus (Marvin, 1998). The N-terminal N1 domain of g3p consists of 68 amino acids and is absolutely essential for *Escherichia coli* infection (Armstrong et al., 1981; Jakes et al., 1988; Stengele et al., 1990; Holliger and Riechmann, 1997; Krebber et al., 1997). The 132 amino acid sized N2 domain, which forms a complex with N1 on the phage (Lubkowski et al., 1998), specifically interacts with the *E. coli* F-pilus (Jakes et al., 1988; Stengele et al., 1990). This pilus interac-

Abbreviations: SIP, selectively infective phage; SAP, selection and amplification of phage; DIRE, direct interaction rescue; g3p, gene-3-protein of filamentous phage; N1, first N-terminal domain of g3p; N2, second N-terminal domain of g3p; CT, C-terminal domain of g3p; hag, hemagglutinin peptide epitope of antibody 17/9

* Corresponding author. Tel.: +41-1-635-5570; fax: +41-1-635-5712; e-mail: plueckthun@biocfebs.unizh.ch

0022-1759/99/$ - see front matter
PII: S0022-1759(99)00143-X



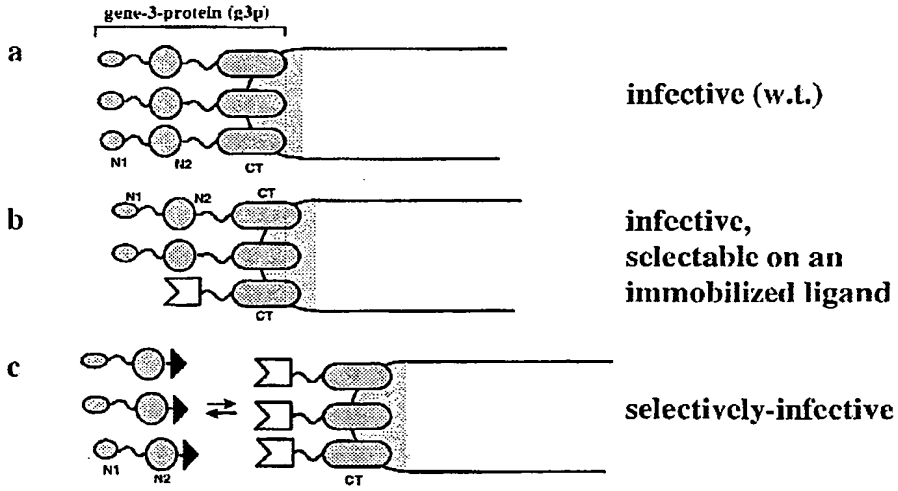


Fig. 1. The modular structure of the minor M13 phage coat protein gene-3-protein (g3p) and its recombinant variations used in phage display and in the SIP technology. For clarity, only three of the probably five copies of g3p are shown. (a) w.t. M13 phage, (b) display phage used in traditional phage display, (c) a variant of a SIP phage, which itself is non-infective and gains infectivity solely in the presence of the adaptor. For details and variations to the particular constructs shown, see text.

tion, however, is not absolutely required for infection, as an alternative, albeit less effective, direct infection pathway exists (Russel et al., 1988; Krebber et al., 1997), which will be described later. The CT domain consists of 149 amino acids (including the C-terminal transmembrane anchor), forms part of the phage coat and is absolutely essential for phage morphogenesis (Nelson et al., 1981; Crissman and Smith, 1984).

In SIP, the basic infectivity of the M13 filamentous phage is destroyed by deleting from the phage genome either the N1 domain or the N1 and N2 domains of the g3p. A peptide or protein library is fused N-terminally to some or all copies of the CT domain or the N2–CT domains of g3p, and no w.t. g3p must be present on the phage. The infectivity of the phage can now only be restored by adding the N1 or the N1–N2 complex, as the N1 domain is absolutely required for infection. These domains are themselves fused or chemically coupled to a ligand which binds to the peptide or protein displayed on the phage. These infectivity restoring molecules will be referred to as the "adaptors", and the consequences of choosing different adaptors, consisting of either N1 or N1–N2, will be discussed later.

There are two routes to selectively restoring the infectivity of the phage: in vivo and in vitro SIP (Fig. 2). For in vitro SIP, both components — the phage displaying the protein and the N1 adaptor or N1–N2 adaptor with the ligand coupled to it — are separately purified and combined in defined amounts in vitro to yield infective phages, provided the ligand binds to the protein. Consequently, the adaptor is encoded on an expression plasmid and the ligand can be either genetically fused to it or, in case of a small organic molecule such as a hapten, chemically coupled to the purified N1–N2 (Gao et al., 1997; Krebber et al., 1997).

In contrast, in the in vivo SIP approach the ligand has to be a protein or peptide genetically fused to N1 or N1–N2, and this fusion protein is encoded on the phage genome. During in vivo phage production, the N1–ligand or N1–N2–ligand adaptor is exported to the bacterial periplasm, while the CT–peptide or CT–protein fusion is also transported to the periplasmic space but remains anchored to the inner membrane through the C-terminal transmembrane helix of CT, before it is incorporated into the budding phage. In case of a tight interaction in the periplasmic space between the polypeptides fused to the adaptor or to

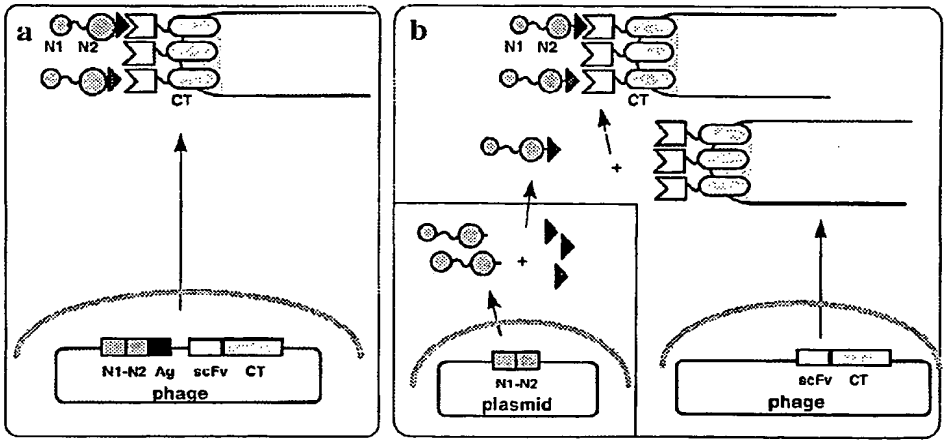


Fig. 2. The principles of in vivo- and in vitro-SIP. The contour of the *E. coli* cells expressing the phage or the adaptor are symbolized by thick grey lines. (a) In the in vivo SIP variant, the phage and the adaptor are produced from the same cell. While the system is drawn with a single replicon, polyphage with copackaged replicons can also be used. In principle, interacting pairs can be coevolved in a library-vs.-library setting, as the genetic information for both libraries is linked in the same phage and propagated in the same cell. (b) For in vitro SIP, the two components are produced separately. Thus, the system can be better controlled. Recombination events between the phage and the adaptor leading to w.t. phages are impossible, and, furthermore, the concentrations of adaptor and phage relative to each other can be controlled in order to drive selection stringently towards higher affinities (see also Fig. 3). However, no coevolution of interacting peptides is possible, as the genetic information of the polypeptide linked to the adaptor is not coupled to the selection process.

the CT domain, respectively, the infectivity of the phage is restored.

The major advantage of SIP in comparison to phage display is the strict coupling of the selection and the infection process, which occur simultaneously. Two further important advantages are apparent for the in vivo SIP approach. First, in identifying an interacting peptide or protein partner to a specific protein, this protein does not have to be first expressed and purified as in phage display. Instead, its DNA is all that is needed, and only very small quantities have to be functionally expressed in the selection system. Nevertheless, it obviously does have to be compatible with transport to and folding in the periplasmic compartment. Second, the in vivo SIP strategy would in principle also be suitable for "library-vs.-library" selections, which are not possible in a direct manner in traditional phage display. However, current limitations in the efficiency of selection, leading to only a limited effective library size, and some unresolved issues in adaptor exchange between phages (see below) have so far not lead to a practical realization of this strategy. On the other hand, progress has been made in developing methods how such "two-dimensional" libraries can in principle be constructed conveniently, as under some circumstances filamentous phages can pack two single-stranded vectors, which may each encode one of the potentially interacting proteins (Rudert et al., 1998).

Since its first proof-of-principle experiments with antibody Fab and scFv fragments as well as with coiled-coil peptides (Dueñas and Borrebaeck, 1994; Gramatikoff et al., 1994; Krebber et al., 1995), progress in understanding the underlying mechanisms has been made, and this knowledge has lead to the construction of improved in vitro and in vivo SIP phage vectors, which have been successfully applied to the selection from various synthetic scFv libraries.

1.2. Structural insight relevant to SIP

New insight has been gained into the structural requirements of fusions to N1 and N2 through the solution of the N1–N2 structure by X-ray crystallography (Lubkowski et al., 1998). Both domains consist mainly of β-sheet and show a striking similarity in their core folds, which suggests an evolutionary

origin by domain duplication. Between the N1 and N2 domains exists a large contact interface formed by two β-strands of N2 that participate in the N1 β-sheet. Nevertheless, there is some flexibility in the relative orientation of N1–N2 (Holliger et al., 1999), and N1 alone has the same structure as in the complex, as determined by NMR (Holliger and Riechmann, 1997). In the infection process, the N2 domain binds to the *E. coli* F-pilus, and while the pilus is "withdrawn", the N1 domain is brought into contact with the C-terminal domain of TolA (Click and Webster, 1997; Riechmann and Holliger, 1997; Click and Webster, 1998; Deng et al., 1999). This interaction appears to be absolutely critical, as no infection is possible at all without either the N1 domain or in the absence of TolA, while the pilus and the N2 domain both merely improve infectivity, but are not indispensable. The crystal structure of the complex of N1 and TolA was solved recently (Lubkowski et al., 1999), and it clearly shows that TolA displaces the N2 domain, which had been proposed from biochemical experiments (Riechmann and Holliger, 1997), even though both bind with very different geometry. Thus, the flexible linkers connecting N1, N2 and CT are an integral part of the rearrangements necessary in the infection process. It is at present not clear what the further fate of the domains is in the infection process nor which further *E. coli* proteins may interact with them. It follows that there may be geometric restrictions in the protein–ligand pairs compatible with SIP, and the affinity threshold (see below) may also be related to the infection mechanism.

2. Recent advances in SIP technology

2.1. Model systems

A thorough study of infection properties of different g3p fusion modules has brought some further understanding of the infection process, especially of the in vitro SIP method (Krebber et al., 1997). In this study, β-lactamase was inserted at different positions within g3p, and also different fusions of a scFv fragment to the phage have been investigated in conjunction with different adaptor constructs. It could be shown that N1 is absolutely required for infection

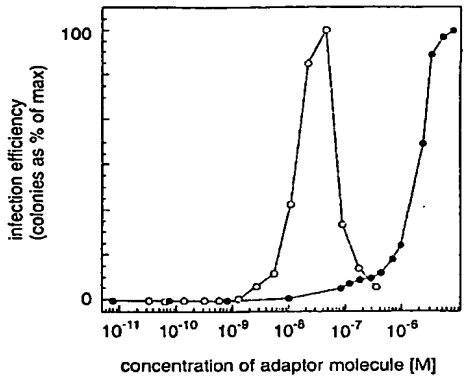


Fig. 3. Different arrangements of the g3p domains for in vitro SIP lead to different infectivity profiles dependent on the adaptor concentration. The N1–N2 adaptor (○) leads to inhibition of infection at higher concentrations and is therefore only suitable for interacting pairs with higher binding constants, and thus ideal for exerting selection pressure towards higher affinities by lowering the adaptor concentration in the infection experiment. The N1 adaptor (●) has to be employed at lower concentrations and might thus be more suitable for interacting pairs with lower affinities. (Figure adapted from Figs. 6 and 7 of Krebber et al. (1997).)

under all circumstances, whereas infection in the absence of N2 is possible, but is dependent on Ca^{2+}. In this case, a pilus-independent infection is made possible by Ca^{2+} disturbing the membrane (Fig. 3).

A library displayed on a SIP phage can be constructed by fusing it N-terminally either to CT or to N2–CT. In general, the N2–CT fusions give higher infectivities (Krebber et al., 1997). Both types can be combined with either an N1–ligand or an N1–N2–ligand, thereby having either zero, one or two copies of N2 in each reassembled g3p. With the N1–N2–ligand adaptor only very low concentrations of adaptor (10^{-8} M) had been necessary for infection in the investigated scFv–hapten system ($K_D = 10^{-10}$ M) (Vaughan et al., 1996), while the same adaptor *inhibits* infection at higher concentrations (10^{-7} M) (Krebber et al., 1997) (Fig. 3, ○). One possible explanation may be that the N1–N2 adaptor binds simultaneously to the pilus and the phage at high concentrations, which is something the N1 adaptor cannot do. Consequently, the N1-N2 adaptor may be very suitable for improving binding constants by SIP by constantly lowering the adaptor concentrations in consecutive rounds. Conversely, the adaptor N1-

ligand gives no inhibition up to adaptor concentrations of about 10^{-6} M, although it may inhibit at even higher concentrations. Higher concentrations are not only possible, but also necessary for infection with the N1 adaptor: In the scFv–hapten system with a K_D of 10^{-10} M an N1 adaptor concentration of 10^{-6} M was required for optimal infection (Krebber et al., 1997) (Fig. 3, ●). Therefore, this adaptor type should be valuable in systems with lower binding constants. It should be pointed out, however, that neither any unequivocal low-affinity system has yet been successfully selected — and, in some cases, the K_D values are simply unknown — nor has this question of the K_D threshold yet been systematically investigated in the various adaptor combinations.

In vitro SIP was also shown to be useful for the selection of catalytic antibodies in a model experiment with a defined antibody (Gao et al., 1997). SIP phages displaying a catalytic antibody scFv fragment fused to CT can be rendered highly infectious, when the catalytic scFv is covalently trapped by a suicide substrate coupled to N1–N2. The coupling chemistry of the suicide inhibitor to N1–N2 was varied, testing the coupling of an engineered cysteine to a substrate containing a maleimide moiety, or interactions between N1–N2 with a His-tag to a substrate coupled to Ni-NTA, or between a N1–N2–streptavidin fusion with a biotinylated substrate. All three coupling procedures lead to selective infectivity, however, for the streptavidin–N1–N2 fusion the infectivities were generally low, which may be due to the tetrameric structure of streptavidin interfering with the infection process.

Finally, in vivo SIP was used in a defined model system for testing a two-vector system for packaging the genetic information for in vivo SIP (Rudert et al., 1998). This would be more convenient for making libraries in both partners at once or for using the same library with many targets without recloning, with a view of "library-vs.-library" screening. This system was tested with the intracellular domain of p75 neurotrophin receptor coupled to the N1–N2 adaptor and an interacting peptide displayed on phage in a CT fusion. Both vectors were packaged in a polyphage after cotransformation, yielding phage particles that were infectious in a cognate pair, but not in a negative control. Infection events could be scored as colonies, when the donor cell streaks were grown on a filter, the phage passed through the filter and infected the recipient on the agar underneath. Polyphage production, which is required in this approach, is generally related to low incorporation of g3p fusion proteins into the phage (note that there is no g3p w.t. in this system), but the exact requirements are not yet clear. The coexistence of a phage and a phagemid genome in the same host require a genetic alteration in the phage genome, termed the "interference resistance" phenotype (Enea and Zinder, 1982).

A different, SIP-related approach of exploiting the selective infectivity of filamentous phages was taken by Sieber et al. (1998). Here, a ribonuclease T1 library was inserted between N2 and CT in the M13 phage. Infectivity was selectively destroyed by protease cleavage, thus selecting for stability and protease resistance, and not, as it is usually done in phage display, for ligand interactions. A similar system was developed by Kristensen and Winter (1998) based on a phagemid/helper phage system using a helper phage with a protease site between N2 and CT. This helper phage will also be useful for normal phage display, as it can help to reduce the background.

2.2. Examples for SIP library selections

In an example of library applications of the in vivo SIP system Gramatikoff et al. (1995) selected ligands to a jun-peptide from a human cDNA library. In contrast to all other examples cited here, they fused the library to the adaptor, while the interacting (constant) jun-peptide was displayed on the phage. No comments on false-positives were given, but in other in vivo SIP projects the adaptor exchange during phage production had led to an uncoupling of phenotype and genotype (S. Spada and D. Christ, unpublished results).

With the in vivo SIP methodology, a larger synthetic library was selected for the most stable scFv structure (Spada et al., 1998). Three successive amino acids in V_L around position 8, which usually is a *cis*-Pro in κ-chains, were randomized by Kunkel mutagenesis in the hemagglutinin (hag) peptide binding scFv 17/9. Only Pro-containing sequences were selected after three rounds of SIP, and it was shown that stability had been the selection criterion rather

than folding yield. An interesting corollary of this result was that the naturally most abundant sequences around position L8 had been selected.

Dueñas et al. (1996) used in vitro SIP in a library setting with a small model Fab library. The selection was shown to require high affinity, and the authors suggested that selection for low and high on- or off-rates could be guided by fine-tuning the selection conditions. Another mini-library of defined point mutants of a fluorescein-binding scFv could be selected by in vitro SIP for a threshold affinity within one round, and for the combined optimum between affinity and the amount of folded and active protein within three rounds (Pedrazzi et al., 1997). In vitro SIP was also employed to select for a useful non-repetitive scFv linker. A linker library was obtained by cloning of a semi-randomized linker cassette into the fluorescein binding scFv FITC-E2 (Vaughan et al., 1996), and SIP-selection yielded all functional scFvs after only a single round (Hennecke et al., 1998).

3. Troubleshooting SIP:pitfalls and countermeasures

While SIP has been shown to be able to select tight binders from libraries in a single round, as well as to be a very powerful technique for the enrichment of the best binder and folder from a library of similar molecules, we have discovered a few pitfalls, which the user needs to be aware of in order to take the appropriate countermeasures for making optimal use of the technology. The selection for tight binding is so powerful that covalent bonds between the adaptor and the phage are strongly selected. This has two consequences, which will both be an issue only in the in vivo SIP method, but not in the in vitro SIP approach. First, there is the danger of picking up mutations in which disulfide linkages are introduced. Second, at some low frequency, w.t.-like phages may appear through a variety of genetic rearrangements, in which the genetic fusion of N1–N2 to CT is restored.

3.1. Selection of spurious disulfide bonds

We have observed the occurrence of unwanted disulfide bonds in two scenarios. In the first, DNA shuffling of a scFv fragment was carried out (S.J. and A.P., unpublished results), which possessed the CDRs of the anti-hag antibody 17/9 (Rini et al., 1992; Schulze-Gahmen et al., 1993; Spada et al., 1998), grafted onto the framework of the scFv B72.3 (Brady et al., 1992; Desplancq et al., 1994). Using in

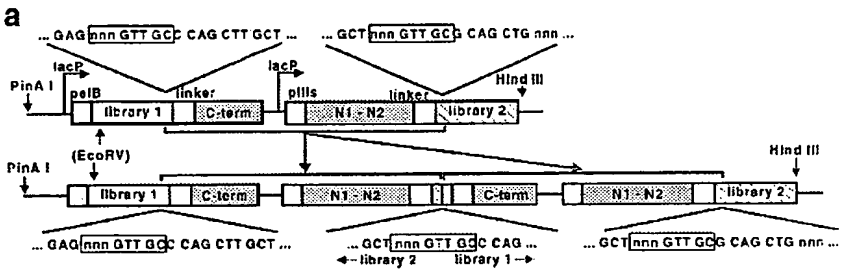
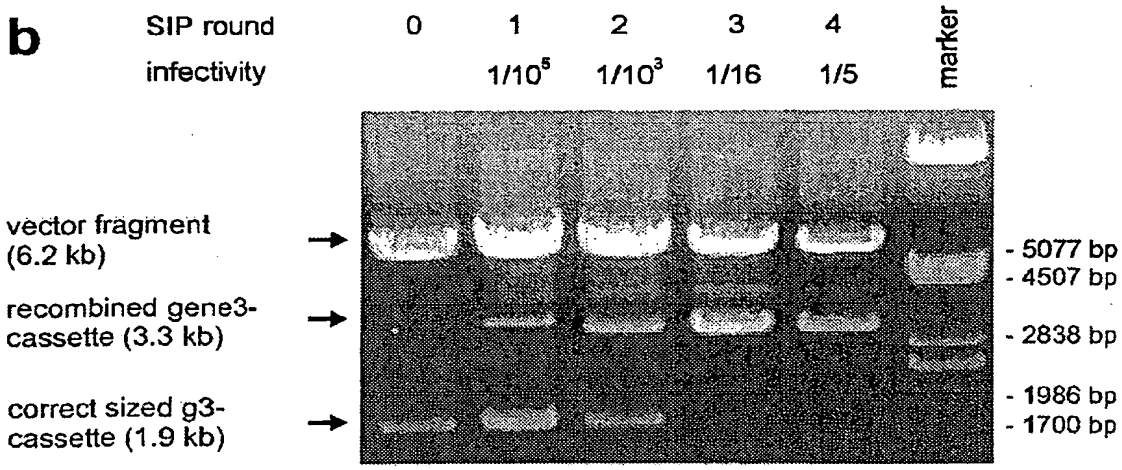


Fig. 4. Detection of recombinations. (a) Proposed molecular mechanism of recombination. The g3 cassette is doubled and recombined at homologous sequence stretches in libraries 1 and 2. (b) Purified phage DNA of jun–fos libraries was digested with the enzymes *Pin*AI and *Hin*dIII, which should result in a vector fragment of 6152 bp and an insert of 1913 bp. For the proposed recombination (a) the insert would be 3329 bp long. Samples are shown with their respective infectivities in the initial library (round 0) and after each SIP round. As molecular weight standard lambda-DNA was digested with *Pst*I. (c) Mini-prep phage DNA of scFv library pools A (phages produced at room temperature) and B (phages produced at 37°C) digested with *Eco*RV/*Hin*dIII after each round of SIP. The expected band at 2174 bp for the insert also occurs in the control phage fB72.3HAG, the library parent. The recombination band at ca. 3600 bp, expected according to a recombination similar to that shown in (a), neither occurs in the parent phage nor in the initial and recloned libraries before SIP selection, but only after the 1st SIP round in both libraries A and B. After recloning, the recombination increases in strength only in library B where the selection pressure is higher than in library A. Nevertheless, the infectivity rises also in the latter library due to disulfide bond formation. Molecular weight standard: λ-DNA cut with *Bst*EII.



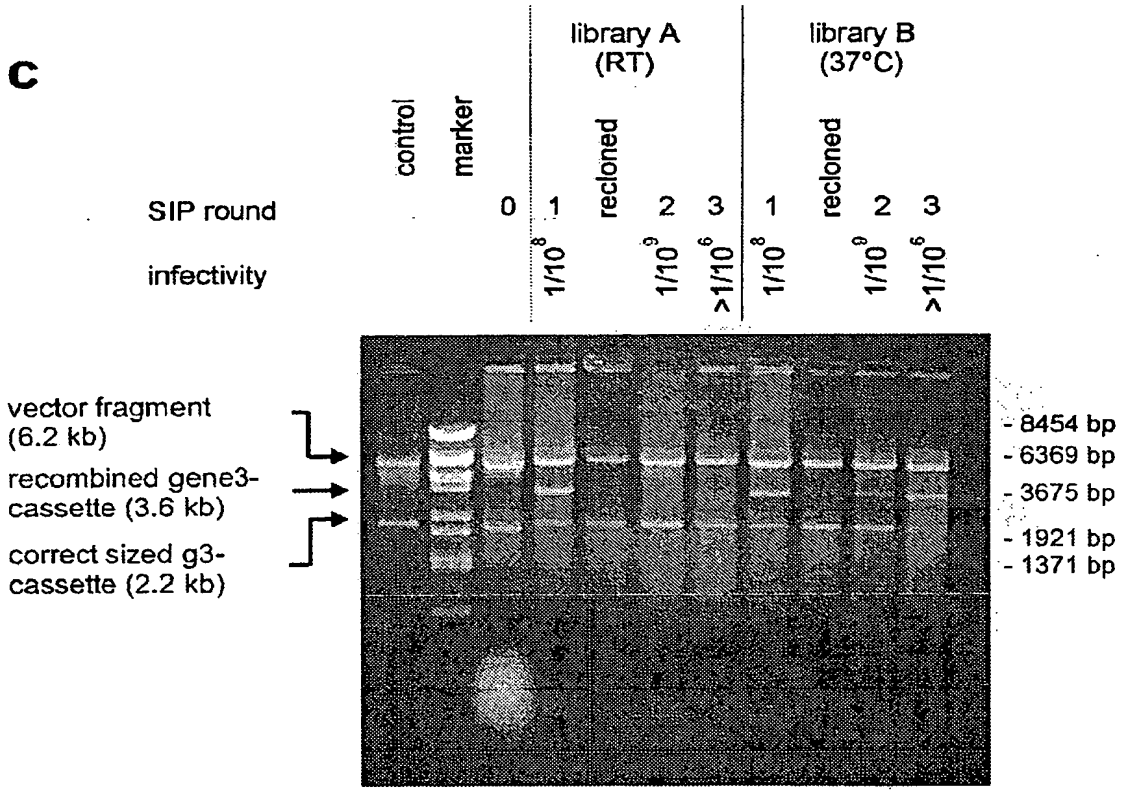


Fig. 4 (continued).

vivo SIP with vectors as described before with the scFv fragment fused to CT (Spada et al., 1998), scFv fragments were enriched carrying unpaired cysteines. In the clones, for which the corresponding N1–N2–hag fragment was sequenced, a frameshift was detected behind N2, leading to a small peptide fusion containing a cysteine. Thus, it appears that a disulfide link was selected which covalently links the adaptor at its C-terminus to the phage-displayed scFv–CT fusion. Indeed, an overnight incubation of these phages with 5 mM DTT at 10°C decreased the infectivity by one order of magnitude, while this treatment decreased the infectivity of w.t. phages and phages displaying a scFv without free cysteines by only twofold.

Similarly, a semisynthetic library of Jun-related peptides was displayed on the phage (K.M.A. and A.P., unpublished results), while a library of Fos-related peptides was fused at the C-terminus of N1–N2. In this library-vs.-library selection, after two rounds, Jun and Fos sequences were enriched, which each contained single cysteines. Interestingly, in the peptide–CT fusions the origin of most of the cysteines were point mutations, most likely introduced in the original PCR-based cassette generation of the library from the long synthetic oligonucleotide. Apparently, the CT domain, which is necessary for the formation of functional phages (see below), largely prevents frameshifts. On the other hand, the peptide fused C-terminally to N1–N2 generated cysteines by frameshifting to other reading frames (see below). For a long synthetic oligonucleotide, 1 bp deletions at a low level are essentially unavoidable and remain present even after purification by polyacrylamide gel electrophoresis.

In the case of the unspecifically interacting disulfide-linked peptides, treatment of the phages with 5 mM DTT at 37°C reduced the cysteine formation by four to eight orders of magnitude, compared to one order of magnitude for wild-type phages, so that further selection rounds were carried out without reappearance of cysteine pairs. Thus, while the use of DTT in experiments where spurious cysteines may occur did reduce the problem, it could not be completely eliminated. It is worth noting that currently all successful in vivo experiments have used defined, high-quality libraries, which were devoid of cysteines (Spada et al., 1998) or have not gone through more than one round (Gramatikoff et al., 1995). It should be stressed that the occurrence of unspecifically paired cysteines is not a problem at all during in vitro SIP, as the adaptor is chemically defined and does not carry a spurious cysteine. Thus, no such problem has been observed during in vitro SIP, even when using multiple rounds (Dueñas et al., 1996; Pedrazzi et al., 1997; Hennecke et al., 1998).

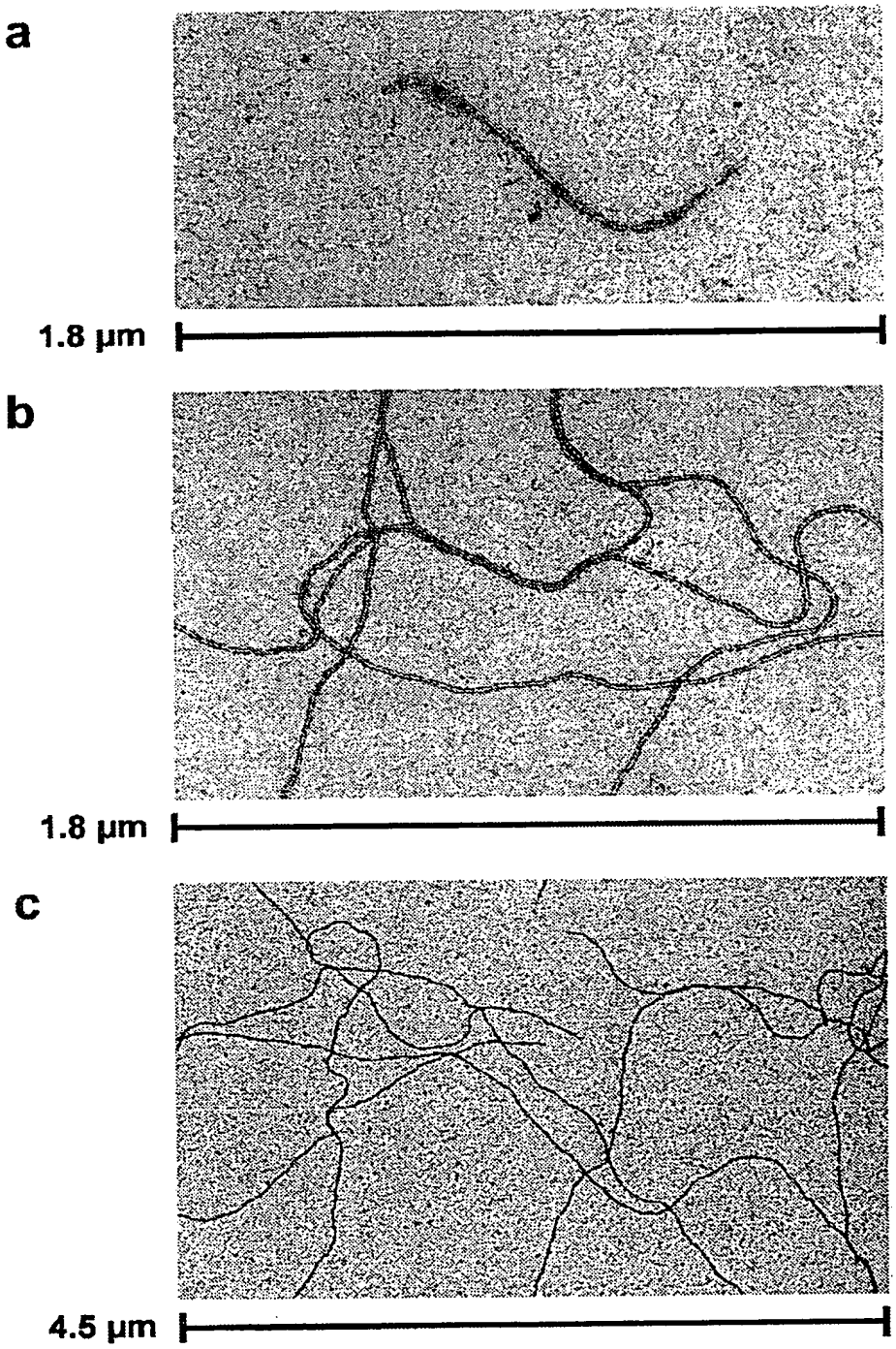
3.2. Genetic recombinations

A second potential pitfall during in vivo SIP is the selection for a genetic recombination which restores some form of N1–N2–CT connection. This genetic rearrangement background was greatly reduced by the use of appropriate vectors (Krebber et al., 1995), but it could not be totally eliminated in some circumstances. Apparently, very short stretches of sequence identity (as short as 8 bp) are sufficient (K.M.A. and A.P., unpublished results), and this cannot always be prevented in library studies, as we have found in two independent library projects using fully randomized synthetic libraries (Fig. 4). However, it is easy to check for and minimize this recombination reaction. Since the size of the restriction fragments encoding the protein–CT and the N1–N2–ligand dramatically changes upon recombination (Fig. 4a), the desired DNA fragment of the original size can be cut from preparative agarose gel electrophoresis gels every few rounds and be recloned into fresh vector, which largely eliminates the problem (Fig. 4b,c). Reducing the phage production temperature also helped to reduce the extent of recombination events (Fig. 4c).

3.3. Frameshifts in the displayed polypeptide lead to functional polyphages

Occasionally, the occurrence of frameshifts, which still allowed the functional production of proliferative phages, have been observed in traditional phage display (Carcamo et al., 1998; Jacobsson and Fryk-

Fig. 5. Electron microscopy of fd phages. The scale is given below each panel. (a) In vivo SIP phage with the g3p divided into two parts as indicated in the scheme of Fig. 1c (b, c) In vivo SIP polyphages in which the CT-domain is out frame, two different scales.


a
1.8 µm
b
1.8 µm
c
4.5 µm

berg, 1998). These frameshifts occurred in the polypeptide N-terminally fused to the CT domain so that the CT domain was out of frame. Also, in SIP experiments, we have observed the occurrence of frameshift variants, which were selected due to unspecific disulfide bond formation as described above. This was seen especially in synthetic libraries generated with long synthetic oligonucleotides, where 1-bp deletion products are common and difficult to separate from full-length oligonucleotides. Although the clones with frameshifts in front of CT should not be functional with the CT domain being out of frame, they were evidently selectable by SIP. As we wished to elucidate the reason for this behavior, we tested whether phages can be formed without the CT domain at all or whether unspecifically "sticky" frameshift polypeptides could replace the CT domain within the phage and be directly incorporated into the phage coat. In such a case, an N1–N2–peptide fusion would have to be able to bind directly to the phage coat.

However, neither the total genetic deletion of the CT domain from the SIP phage nor the replacement of the CT domain with the short frameshifted polypeptides in w.t. phages lead to any detectable infectivity using 10^{11} phages for infection. In contrast, the infectivities of these selected clones, possessing unpaired cysteines and frameshifts in the genetic presence of a CT domain, lead to a relatively high SIP infectivity of about $1/10^5$ phages. Therefore, the CT domain must be present on the protein level in the genetically frameshifted SIP phages, possibly through a second, *translational* frameshift, bringing the CT domain back into the right frame.

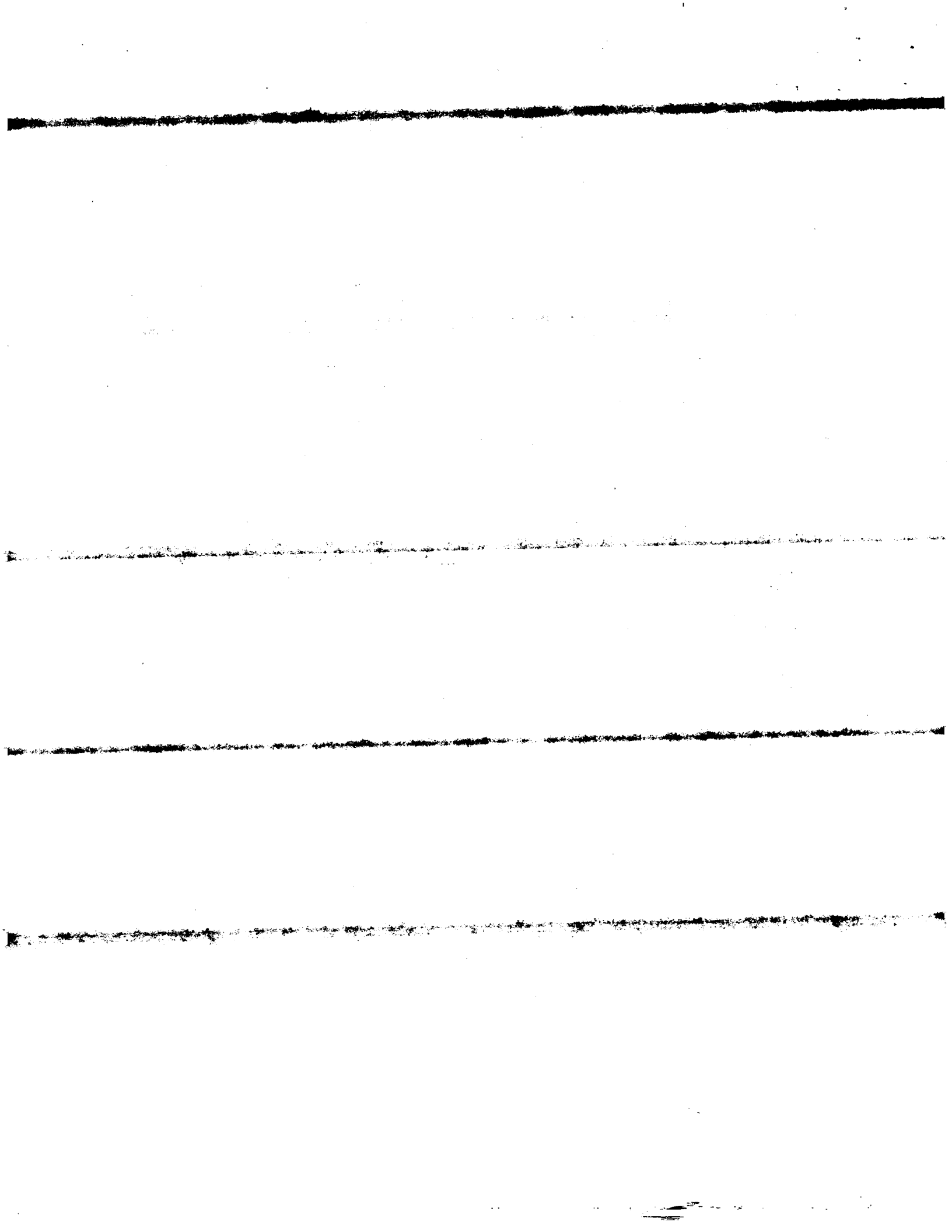
In fact, two frameshifted clones investigated more closely possessed two subsequent rare arginine codons, AGG, which are known to promote translational frameshifts (Spanjaard and van Duin, 1988). These clones were shown to produce polyphages, as detected by electron microscopy (Fig. 5). In conclusion, the frameshift variants occasionally observed to be selectable by traditional phage display (Carcamo et al., 1998; Jacobsson and Frykberg, 1998) or in SIP do possess a functional CT domain on the protein level, but the phage-producing cell makes so little of it, due to the rare events of translational frameshifts, that polyphages — which are infective — are produced. As long as the adaptor is covalently linked (via the spurious disulfide bonds), these phage can be selected. Importantly, there is apparently no danger of "direct" binding of the adaptor to the phage via non-specific interactions, as the CT domain is absolutely required for functionality.

4. Conclusions

While the in vivo SIP technology is especially convenient, as no protein at all needs to be expressed and purified for the selection of binding partners, it is important to understand the potential side reactions which can result in false positives: spurious cysteines, leading to covalently disulfide-linked adaptor–phage complexes, and rare genetic recombinations which regenerate N1–N2–CT rearrangements. Recombination events can be efficiently eliminated by recloning of the correct-sized g3p cassette. While DTT incubations can reduce much of the disulfide coupling, it does not reduce the background to zero. Furthermore, *genetic* frameshifts leading to nonsense-polypeptides and spurious cysteines are not necessarily strictly selected against, probably due to a low frequency of *translational* frameshifts promoted by certain sequences, which bring the CT domain back into frame so that functional phages can be produced in spite of frameshifts. Therefore, caution must be exercised in applying in vivo SIP to libraries obtained from error-prone PCR, DNA shuffling or very long oligonucleotides, and especially from cDNA.

In contrast, very encouraging results have been obtained with defined in vivo SIP libraries, such as the randomization of a short stretch of a V_L domain (Spada et al., 1998). In this case, the library was well defined and free of spurious cysteine codons at the level required. It should be stressed again that none of the problems occur during in vitro SIP, even after several rounds, and a number of libraries have been successfully screened (see below).

The potential advantage of in vitro SIP has been the very low background, at least under all conditions tested, which allowed functional molecules to be selected after only one single round of SIP selection (Krebber et al., 1997; Hennecke et al., 1998). While in traditional phage display enrichment factors of $10–10^4$ per round are normal (Winter et al., 1994), enrichment factors of $10^5–10^6$ per round can



be easily achieved using SIP (Dueñas and Borrebaeck, 1994; Dueñas et al., 1996). Additionally, a well-defined library permits a convenient enrichment of molecules with even small advantages in molecular properties (Pedrazzi et al., 1997; Spada et al., 1998). Thus, we see in vivo SIP, at the current level of understanding, mostly as a technology for molecular improvement, and less one of initial screening of large libraries, except in such cases where very high affinities will be present.

In vitro SIP is far more resistant to spurious genetic alterations, as the N1–N2–ligand or the N1–ligand adaptor is constant, since its genetic information is not amplified together and coevolved with the partner displayed on the phage. Several library experiments have been successfully carried out, including both defined mutant libraries (Dueñas et al., 1996; Pedrazzi et al., 1997) and partially randomized libraries (Hennecke et al., 1998). For in vitro SIP, the high required (affinities of at least 10^{-9} M) have been well documented (Dueñas et al., 1996; Krebber et al., 1997; Pedrazzi et al., 1997), even though the exact number may depend on the molecular system in question, and for some of the model systems reported, the K_D is not known.

SIP is a powerful strategy to select for protein–ligand interactions as well as for other desired features as protein folding and stability. Moreover, the threshold for the phage to infect seems to be so high that the selection pressure for the very best restored g3p is enormous, meaning that excellent binding or even covalent linkage of the N-terminal domains and the C-terminal domain is strongly favored within the selection process. Provided that artifacts can be controlled, by using high-quality libraries and/or in vitro SIP, this is a big advantage of SIP compared to traditional phage display, as SIP can within very short time and with minimal effort select for the best binders and even discriminate subtle differences (Pedrazzi et al., 1997; Spada et al., 1998). In cases where selection for a covalent binder is actually *desired*, like in the trapping of catalytic antibodies by suicide inhibitors (Gao et al., 1997), for the selection of interacting pairs with extremely low dissociation constants or by separating proteolytically cleaved proteins from intact ones (Kristensen and Winter, 1998; Sieber et al., 1998), this technology is very attractive because of its speed and selection power and because no dissociation of the tightly or covalently interacting pair is required in SIP.

In summary, SIP is an extremely rapid and powerful selection alternative to conventional phage display. The applicability of in vivo SIP can be extenuated to arbitrarily randomized libraries and libraries with low initial infectivity only when special precautions are taken to guide selection towards non-covalent interactions and to prevent the selection of genetically or chemically restored w.t. phages. As these precautions do not completely suppress undesired covalent variants of g3p, the in vivo SIP methodology is more suitable for libraries made by controlled mutagenesis and with sufficiently interacting pairs initially present. However, if only one library is to be screened against a constant partner, the in vitro SIP variant of the SIP methodology should be the method of choice, as it is more robust against spontaneous genetic changes within the phage. Recombination of w.t. g3p and spontaneous mutations towards cysteines on the N1–N2 adaptor are not possible in in vitro SIP, as the N-terminal domains are not genetically linked in the selection. The future development and extension of SIP, however, will clearly require a more detailed mechanistic understanding of the phage infection process.

Acknowledgements

We thank René Hermann (ETH Zürich) for the electron microscopy experiments.

References

Armstrong, J., Perham, R.N., Walker, J.E., 1981. Domain structure of bacteriophage fd adsorption protein. FEBS Lett. 135, 167.

Brady, R.L., Edwards, D.J., Hubbard, R.E., Jiang, J.S., Lange, G., Roberts, S.M., Todd, R.J., Adair, J.R., Emtage, J.S., King, D.J., Low, D.C., 1992. Crystal structure of a chimeric Fab' fragment of an antibody binding tumour cells. J. Mol. Biol. 227, 253.

Carcamo, J., Ravera, M.W., Brissette, R., Dedova, O., Beasley, J.R., Alam-Moghe, A., Wan, C., Blume, A., Mandecki, W., 1998. Unexpected frameshifts from gene to expressed protein in a phage-displayed peptide library. Proc. Natl. Acad. Sci. U.S.A. 95, 11146.

Click, E.M., Webster, R.E., 1997. Filamentous phage infection: required interactions with the TolA protein. J. Bacteriol. 179, 6464.

Click, E.M., Webster, R.E., 1998. The TolQRA proteins are required for membrane insertion of the major capsid protein of the filamentous phage f1 during infection. J. Bacteriol. 180, 1723.

Crissman, J.W., Smith, G.P., 1984. Gene-III protein of filamentous phages: evidence for a carboxyl-terminal domain with a role in morphogenesis. Virology 132, 445.

Deng, L.W., Malik, P., Perham, R.N., 1999. Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. Virology 253, 271.

Desplancq, D., King, D.J., Lawson, A.D., Mountain, A., 1994. Multimerization behaviour of single chain Fv variants for the tumour-binding antibody B72.3. Protein Eng. 7, 1027.

Dueñas, M., Borrebaeck, C.A., 1994. Clonal selection and amplification of phage displayed antibodies by linking antigen recognition and phage replication. Biotechnology 12, 999.

Dueñas, M., Malmborg, A.C., Casalvilla, R., Ohlin, M., Borrebaeck, C.A., 1996. Selection of phage displayed antibodies based on kinetic constants. Mol. Immunol. 33, 279.

Enea, V., Zinder, N.D., 1982. Interference resistant mutants of phage f1. Virology 122, 222.

Gao, C., Lin, C.H., Lo, C.H.L., Mao, S., Wirsching, P., Lerner, R.A., Janda, K.D., 1997. Making chemistry selectable by linking it to infectivity. Proc. Natl. Acad. Sci. U.S.A. 94, 11777.

Gramatikoff, K., Georgiev, O., Schaffner, W., 1994. Direct interaction rescue, a novel filamentous phage technique to study protein–protein interactions. Nucleic Acids Res. 22, 5761.

Gramatikoff, K., Schaffner, W., Georgiev, O., 1995. The leucine zipper of c-Jun binds to ribosomal protein L18a: a role in Jun protein regulation? Biol. Chem. 376, 321.

Hennecke, F., Krebber, C., Plückthun, A., 1998. Non-repetitive single-chain Fv linkers selected by selectively infective phage (SIP) technology. Protein Eng. 11, 405.

Holliger, P., Riechmann, L., 1997. A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5, 265.

Holliger, P., Riechmann, L., Williams, R.L., 1999. Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Å: evidence for conformational lability. J. Mol. Biol., in press.

Jacobsson, K., Frykberg, L., 1998. Gene VIII-based, phage-display vectors for selection against complex mixtures of ligands. Biotechniques 24, 294.

Jakes, K.S., Davis, N.G., Zinder, N.D., 1988. A hybrid toxin from bacteriophage f1 attachment protein and colicin E3 has altered cell receptor specificity. J. Bacteriol. 170, 4231.

Krebber, C., Spada, S., Desplancq, D., Plückthun, A., 1995. Co-selection of cognate antibody–antigen pairs by selectively infective phages. FEBS Lett. 377, 227.

Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L., Plückthun, A., 1997. Selectively infective phage (SIP): a mechanistic dissection of a novel in vivo selection for protein–ligand interactions. J. Mol. Biol. 268, 607.

Kristensen, P., Winter, G., 1998. Proteolytic selection for protein folding using filamentous bacteriophages. Folding Des. 3, 321.

Lubkowski, J., Hennecke, F., Plückthun, A., Wlodawer, A., 1998. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. Nat. Struct. Biol. 5, 140.

Lubkowski, J., Hennecke, F., Plückthun, A., Wlodawer, A., 1999. Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. Structure 7, 711.

Marvin, D.A., 1998. Filamentous phage structure, infection and assembly. Curr. Opin. Struct. Biol. 8, 150–158.

Nelson, F.K., Friedman, S.M., Smith, G.P., 1981. Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III. Virology 108, 338.

Pedrazzi, G., Schwesinger, F., Honegger, A., Krebber, C., Plückthun, A., 1997. Affinity and folding properties both influence the selection of antibodies with the selectively infective phage (SIP) methodology. FEBS Lett. 415, 289.

Riechmann, L., Holliger, P., 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. Cell 90, 351.

Rini, J.M., Schulze-Gahmen, U., Wilson, I.A., 1992. Structural evidence for induced fit as a mechanism for antibody–antigen recognition. Science 255, 959.

Rudert, F., Woltering, C., Frisch, C., Rottenberger, C., Ilag, L.L., 1998. A phage-based system to select multiple protein–protein interactions simultaneously from combinatorial libraries. FEBS Lett. 440, 135.

Russel, M., Whirlow, H., Sun, T.P., Webster, R.E., 1988. Low-frequency infection of F-bacteria by transducing particles of filamentous bacteriophages. J. Bacteriol. 170, 5312.

Schulze-Gahmen, U., Rini, J.M., Wilson, I.A., 1993. Detailed analysis of the free and bound conformations of an antibody. X-ray structures of Fab 17/9 and three different Fab-peptide complexes. J. Mol. Biol. 234, 1098.

Sieber, V., Plückthun, A., Schmid, F.X., 1998. Selecting proteins with improved stability by a phage-based method. Nat. Biotechnol. 16, 955.

Spada, S., Honegger, A., Plückthun, A., 1998. Reproducing the natural evolution of protein structural features with the selectively infective phage (SIP) technology. The kink in the first strand of antibody kappa domains. J. Mol. Biol. 283, 395.

Spanjaard, R.A., van Duin, J., 1988. Translation of the sequence AGG–AGG yields 50% ribosomal frameshift. Proc. Natl. Acad. Sci. U.S.A. 85, 7967.

Stengele, I., Bross, P., Garces, X., Giray, J., Rasched, I., 1990. Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. J. Mol. Biol. 212, 143.

Vaughan, T.J., Williams, A.J., Pritchard, K., Osbourn, J.K., Pope, A.R., Earnshaw, J.C., McCafferty, J., Hodits, R.A., Wilton, J., Johnson, K.S., 1996. Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. Nat. Biotechnol. 14, 309.

Winter, G., Griffiths, A.D., Hawkins, R.E., Hoogenboom, H.R., 1994. Making antibodies by phage display technology. Annu. Rev. Immunol. 12, 433.